ACOUSTIC CHARACTERISTICS OF (HINDI) VOWEL SOUNDS IN DIFFERENT CONTEXTS

DISSERTATION

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF

Master of Philosophy

IN

PHYSICS

BY

ISRAR KHAN

DEPARTMENT OF PHYSICS
ALIGARH MUSLIM UNIVERSITY
ALIGARH (INDIA)

1988

*
***
****

Dedicated to my parents

&

Loving Brothers

****
***
*

ACKNOWLEDGEMENT

I wish to express my gratitude to shri S. K. Gupta, Lecturer, Department of Physics, for his invaluable guidance throughout the investigation.

I sincerely thanks Prof. Mohd. Shafi, Chairman Department of Physics, A. M. U., Aligarh for permitting me to conduct this study.

My regards to Dr. S. S. Agrawal, Scientist CEERI centre, New Delhi and Mr. A. M. Ansari, Scientist CEERI centre, New Delhi, for their invaluable help and support.

The main credit goes to Dr. Arshad Ahmad, Reader, Department of Physics, A. M. U., for their invaluable help, discussion and guidance.

Last but not the least, I am thankful to the friends and the staff who rendered a helping hand in this work.

ISRAR KHAN

TABLE OF CONTENTS

Chapter 1

G E N E R A L I N T R O D U C T I O N

INTRODUCTION:-

One of the most common mode of communication between human beings is speech. Man is unique in his ability to transmit information with his voice. Only man has developed the vocal means for coding and conveying information beyond a rudimentary stage. Acoustically speaking, speech is a sequence of sound signal, the fundamental frequency , intensity and spectral distributions of which vary from instant to instant. Man from his past experience sends neural signals from brain which actuate the vocal apparatus. Speech sound through mouth and nostrils radiate into the air which are detected by ear. These waves vibrate the eardrum. The vibrational energy is converted into nerve impulses which are sent to brain through auditory nerve for interpretation. Throughout his evolution man has naturally adopted to this communication system. Now an experienced listener encode not only thoughts but talkers emotions, his identity and his very physical state of being. He can also localise sounds and can direct his attention among several talkers. According to **FANT**, " speech is a feedback-mediated, output-oriented integrations of movements in space and time executed by a complex of excitory and inhibitary muscle activities ". Auditory, tactile and proprioceptive feedback loops appear to operate according to a principle of flexibility ensuring the most adequate output in any contextual frame.

Speech as a mode and means of communication is far

ahead of even the advance high technology media communication developed to date. The need of speech as means of communication can best be appreciated by those who can not speak at all and by those who have problem in the communication with others.

Contemporary studies in the field of speech analysis are motivated by the ultimate goal of building automata through which verbal communication between man and machine could be realised. To extend the man 's capibilities and to increese the productivity of human beings, utilization of speech for communication between man and machine has been significant requisite and the main motivation factor developing speech interactive systems. The basic mechanism of speech communication with machine is to functionally duplicate the behavior of human communication link. In order to accomplish this above goal and that of speech perception the principal effort have been directed toward the building of three types of machines:

(1) Machines that can encode linguistic symbols into some sequence of speech sounds that human listener can understand. These machines are called speech synthesizers.

(2) Machines that can decode the acoustic speech signal into its printed equivalent in the form of some sequence of recognizable symbols or printed words. These machines are often referred to as speech recognisers.

(3) Speaker identification / verification : Other advantages.

The synthesis of speech is at present

successfully approached by various methods, speech recognition holds still a variety of unsolved but interesting problems. What makes the problem of decoding speech particularly difficult is the extremely ambiguous nature of the speech code. There are considerable differences in the acoustic speech signals of two utterances of the same word by the same speaker. Matters are complicated a great deal more by the variations of the speech signals from one speeker to another as the result of different speech habits, accent, pitch, stress, etc., of the speakers.

The problem of decoding the speech signal represents a major challenge in achieving an efficient natural communication link between man and machines. In spite of considerable research efforts in this field the results are very modest and there remain a number of fundamental questions which have yet to be answered. The most important of these concern the following issues:

(1) The choice of an adequate set of measurements,

(2) The definition of the basic linguistic units of the machine,

(3) The machine representation of utterances and words, and

(4) The time segmentation of the speech signal.

The basic linguistic units of the decoder are "machine events" as opposed to "phoneme" or other linguistic units generally in use.

The common feature appears to be that the preprocessing of the speech signal is carried out by passing

it through a bank of analog filters with smoothing on the rectified output of each channel. The resulting slowly varying output of each channel is proportional to the square root of the power in that channel. The ensemble of these outputs is generally referred to as the "short time-spectrum". An important observation was that the short-time spectrum of the vowel sounds exihibits peaks at various frequencies, which are known today as "formant frequencies".

The scope of these efforts was limited to the recognition of vowels in some well-defined context and that of the ten spoken digits pronounced carefully under laboratory conditions. The difficulties should not be underestimated. Speech is an extremely irregular signaling system. Everyone's voice is different and so is the voice of a given talker in different situations. These considerations alone should be sufficient to show that the code of speech is not at all simple to break. The synthesis of sounds generated by human vocal tract or by musical instruments proved to be a task for easier than the recognition problem. While the highly irregular nature of speech is a drawback in recognition, the permissible irregularities are in fact advantageous in synthesis.

It is obvious that speech has many complex sounds. To analyze these sounds the use of contextual information is helpful. The contextual informations about the sounds are used by the man to recognize speech. It follows therefore that for recognition higher sources of knowledge are needed. The varios type of knowledge sources which are required to

operate at various levels could be listed as:

(1) The characteristics of speech sounds.

(2) Variability in pronounciation.

(3) Stress patterns.

(4) Sound patterns of words and dictionary (Lexicon).

(5) Grammatical structure of language.

(6) Meaning of words and sentences.

(7) Context of conversation.

It is obvious that only after introducing these sources of knowledge we will be able to recognize more correctly the phonemes, syllables and words of the language in continuous speech.

Speech sounds are classified as: (1) vowels, and (2) consonants. Most of the consonants can not be pronounced without vowels. Therefore, first of all we got interested in study the different aspects of vowels.

Chapter 2

F E A T U R E S O F S P E E C H S O U N D S

SPEECH PRODUCTION:-

The machinery involved in speech production is shown schematically in Fig. 2.1. The primary fuction of inhalation is accomplished by expanding the rib cage, reducing the air pressure in the lungs, and drawing air into the lungs via nostrils, nasal cavity,velum port and trachea (windpipe). Air is normally expelled by the same route. The vocal tract proper is an acoustical tube which is nonuniform in cross-sectional area. It is terminated by the lips at one end and by the vocal cord constriction at the top of the trachea at the other end. The cross-sectional area may be varied by movements of articulators, the lips, jaw, tongue, and velum. The nasal tract constitutes an additional path for sound radiation. It begins at the velum and terminates at the nostrils. Acoustic coupling between the nasal and vocal tract is controlled by the size of the opening at the velum.

In a study to estimate maximal articulatory rates, **HUDGINS and STETSON (1937)** asked talkers to repeat simple syllables as rapidly as possible in rhythmic groups. Their results showed that 8.2 syllables per second could be produced with the tip of the tongue, 7.3 with the jaw, 7.1 with the back of the tongue, and 6.7 with the velum and with the lips. These maximal rates can not be maintained indefinitely because speech is a function of the respiratory system. So, these rates depends upon breathing and articulatory positions.

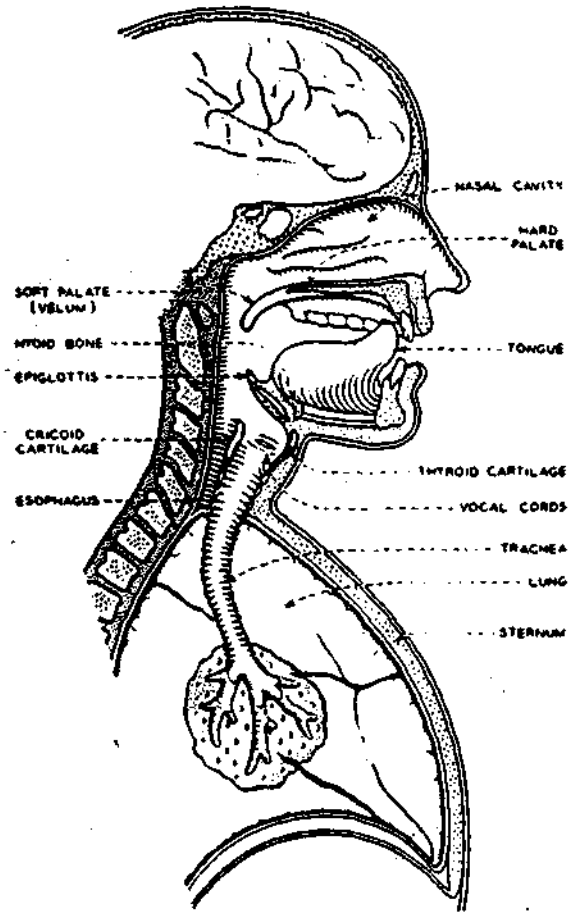


Fig. 2.1. Schematic diagram of the human vocal mechanism

CLASSIFICATION OF SPEECH SOUNDS:-

Classification of speech sounds is accomplished according to their place of production. Phonetician have found this method convenient to indicate the characteristics of sounds. This classification method will be employed in the discussion of speech sounds as:

(1) Consonants, and

2) Vowels.

CONSONANTS:-

The consonants constitute those sounds which are not exclusively voiced and mouth-radiated from a relatively stable vocal configuration. They generally are characterized by greater tract constrictions than the vowels. They may be excited or radiated differently, or both. Consonants may be classified as:

FRICATIVE CONSONANTS: Fricatives are produced from an incoherent noise excitation of the vocal tract. The noise is generated by turbulent air flow at some point of constriction. Radiation of fricatives normally occurs from the mouth. If the vocal cord source operates in conjuction with the noise source, the fricative is a voiced fricative. If only the noise source is used, the fricative is unvoiced. Both voiced and unvoiced fricatives are continuant suunds. Hindi has only unvoiced fricatives.

STOP CONSONANTS: TO produce stop sounds a complete closure is formed at some point in the vocal tract. The lungs build up pressure and the pressure is suddenly released by an abrupt motion of the articulators. The expansion and

aspiration of air help to characterize the stops.

NASAL ONSONANTS: The nasal consonants are normally excited by the vocal cords and hence are voiced. A complete closure is made towards the front of the vocal tract either by the lips, or by the tongue at the gum ridge, or by the tongue at the hard or soft palate. The velum is opened wide and the nasal tract provides the main sound transmission channel.

GLIDES AND SEMIVOWELS: Two small group of consonants contain sounds that greatly resemble from vowels. The glides are dynamic sounds, invariable precede a vowel, and exhibit movement toward the vowel. The semivowels are continuants in which the oral channel is more constricted than in most vowels, and the tongue tip is not down.

VOWELS:-

The vowel sounds of speech are normally produced by vocal cord (or voiced) excitation of the tract. In normal articulation, the tract is maintained in a relatively stable configuration during most of the sounds. The vowels are further characterized by nasal coupling, and by radiation only from the mouth.

If the nasal tract is effectively coupled to the vocal tract during the production of a vowel, the vowel becomes nasalized. The ten most frequent vowels of Hindi speech are classified according to the tongue-hump-position/ degree of constriction scheme, they may be arranged as shown in **Table - I.** Along with each vowel is shown its Hindi orthographic representation. The approximate articulatory

configurations for the production of these sounds are shown qualitatively by the vocal tract profiles in **Fig. 2.2 (POTTER, KOPP and GREEN).**

TABLE-I

Degree of constriction	Tongue-hump-position	
	front	back
High	/i/ ई	/u/ ऊ
	/I/ इ	/U/ उ
Medium	/e/ ऐ	/O/ ओ
	/ɛ/ ए	/ɔ/ औ
Low	/a/ आ	/ʌ/ अ

FEATURES OF VOWEL SOUNDS:-

Vowel sounds have various features. The remaining portion of this chapter will covered some parts of its features only.

Phonetic qualities are defined in terms of their assumed production by the vocal organ's **(BLOCH, 1950). BLOCH and TRAGER (1942)** gives 42 different vowel symbols and recognize the possibility of having to use additional diacritics. **RUSSELL (1928)**, in his work on the relation between vowel quality and articulation, denotes the quality of the vowel by reference to key-words. But **PETERSON (1952)** ensured that the vowels pronounced by diffent subjects were

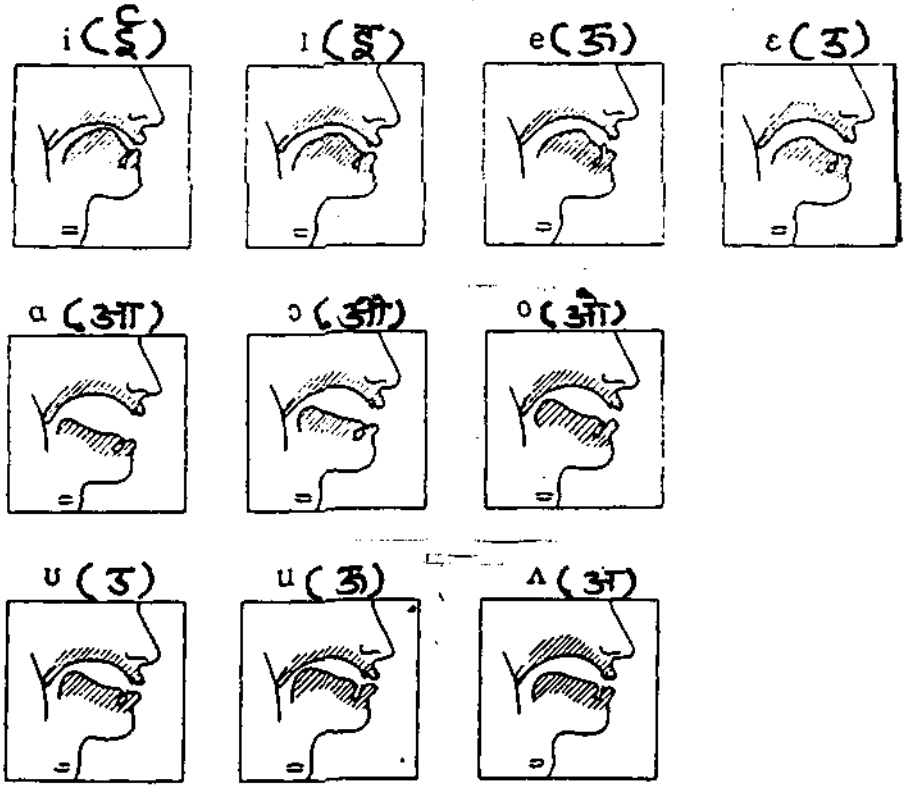


Fig. 22. Schematic vocal tract profiles for the production of Hindi vowels.
(Adapted from POTTER, KOPP and GREEN)

infact phonetically identical.

A phonetician has to be capable of distinguishing many more vowel qualities than there are in his own speech. The listenig phonetician assesses the sound, but he does not consider the sound as a whole. He focuses his attention on certain features of the auditory sensation by which he can determine the quality of the vowel. Each sounds can be classified according to three seperate factors: pitch, loudness, and quality. When two vowels are different usually do not mean any thing about their pitch and loudness nor about their personal quality but only they differ one aspect of their quality which is termed as phonetic quality. The problem can be represented by means of diagrams in two dimension, differences in personal quality along one axis and variations in phonetic quality along the other.

The peculiar position of speech sounds is due to their being habitually assessed as part of a means of communication. Every speaker has learnt to seperate personal quality from phonetic quality as a result of his constant experience. It is assumed that personal quality can be identified with vocal cord quality and phonetic quality with articulatory quality. According to **PETERSON**, the phonetic value of speech sound is independent of language and meaning **(PETERSON, 1952)**. Personal quality depends not only on the mode of vibration of the vocal cords, but also on some acoustic features due to the positions of the articulators.

There are two main lines of vowel development

involved: the articulatory and the acoustic. The first writers to attempt to describe the position of the vocal organs during the pronounciation of vowel sounds was **ROBERT ROBINSON (1617).** He gave a schematic diagram of the articulators. (The next important articulatory description of vowel quality is due to **WALLIS (1653).**) **WILLIAM HOLDER (1669)** used a unidimensional system for classifying vowel quality and suggested taking care of extra factors into account. **A. M. BELL (1867)** had used the current form of description in which vowels were classified in terms of two series "labials" and " linguals" plus an intermediate series "labiolingual". In **BELL's** classification scheme 36 vowel qualities were specified.

SWEET (189Ø) modified and elaborated **BELL's** system so that he was able to specify 72 vowel qualities. All these vowels are given articulatory descriptions. **RUSSEL (1928)** gave the articulatory description and seems that some of auditory impressions of vowel quality may be more simply correlated with acoustic measurements rather than with articulatory data.

The first major contribution to the acoustics of speech was made by **WILLIS (1829).** From his experiment, he came to the conclusion that there were two acoustic features for each vowel sound, the pitch and its own characteristic note. **JOOS (1948),** and **POTTER** and **PETERSON (1948)** acknowledge their recognition of the importance of formant frequencies due to development of sound spectrograph. Previously all the observations of formant frequencies had

been based on measurements of the acoustic characteristics of the vowels of a few individuals. But now, after analysis of a large number of vowels, it became apparent that vowels which were considered to be phonetically equivalent did not necessarily have the same acoustic characteristics.

PETERSON's studies of vowels was one of the first to attempt to explain the precise relationship between the phonetic quality of a vowel and its acoustic characterics. So far one is unable to use spectrographic data or other records of the acoustic characteristics of a vowel to give as exact a specification of its phonetic quality as can be given by a well trained phonetician. The spectrograms can often use to assess the relative phonetic quality of two vowels spoken by the same speaker but can not make any precise statements about the relative phonetic quality of two vowels spoken by different speakers.

JOOS (1948) came to the conclusion that the phonetic quality of vowel depends on the relationship between the formant frequencies for that vowel and the formant frequencies of other vowels pronounced by that speaker. A necessary part of **JOOS's** theory is that whatever a listener to speech has to identify a vowel without the benifit of any cues from the context, the utilizes whatever knowledge he has of the speaker's formant frequencies in other words. On this theory the phonetic value of a vowel depends on the way in which its acoustic structure fits into the pattern formed by the acoustic structure of other vowels produced by the same

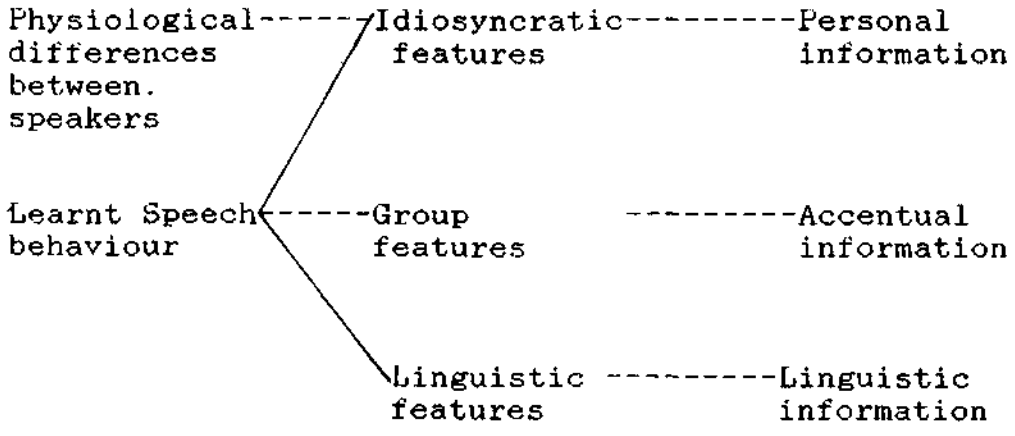
speaker.

One of the well established standards in vowels are a sets of vowels produced by a reliable phonetician trained by **DANIEL JONES** as these will be identical in phonetic quality and is termed as sets of cardinal vowels. These cardinal vowel sounds are very useful in exeperimental investigations of phonetic quality. Therefore, the only vowel sounds which are always said to be identical in quality irrespective of the speaker are cardinal vowels. The Haskins laboratory group, says (one of their speaker) that either the vowel qualities of the cardinal vowels will depend on the native language of the speaker, or they are intended to occupy areas rather than specific points in the vowel continuum. **LADEFOGED's** point of view is that neither of these implications is correct. Since the cardinal vowels are not sets of sounds produced according to certain specifications, but sounds judged by competent observers to have certain phonetic qualities.

PETERSON (1952) has said that the fundamental phonetic parameters should have the same value when the vowel value is the same, regardless of the type of speakers. But **PETERSON** further said that no simple principle has yet been found for obtaining the same parameter values when the same vowel value is pronounced by different types of speakers. Although cardinal vowels are defined in terms of a number of properties one of which is auditory equidistance. When a speaker pronounces a set of cardinal vowels, he attempts to imitate a series of sounds which he has learnt

through aural instruction. Therefore, auditory equidistance may be a property associated to cardinal vowels solely by their originator.

We now discuss about the nature of quality differences, and consider the kinds of information that are conveyed by speech. This information is to be of three kinds. Firstly, when we listen to a person talking we can receive information about what he is saying. Secondly, in addition to the information we receive as a result of considering an utterance in terms of a linguistic system conveying lexical and grammatical information. We also receive information of a different kind about the general background of the speaker. This kind of information may be termed accentual. Thirdly, there is the kind of information conveyed by the idiosyncratic features of a person's speech. The relations between these three kinds of information are summarized as below. Some efforts had been made to arrange experimental situations which will elicit responses with respect to each of these three kinds of information



Whenever a listener to speech has to assess the phonetic quality of a vowel, he utilizes whatever knowledge he has of the speaker's formant frequencies. The hypothesis is that vowels are assessed atleast partly in terms of the way in which their acoustc structure fits into the pattern of sounds which the listener has been able to observe or considers to be probable. Two sounds can convey different linguistic information only if they have different phonetic qualities.

The linguistic information conveyed by a given vowel is partly dependent on the relations between the frequencies of its formants and the frequencies of the formants of other vowels occuring in the same auditory context. Two points may be noted concerning the sociolinguistic and personal information conveyed by vowels. Firstly, there do not appear any differences in the sociolinguistic information conveyed by the different versions of the auditory sentence. Therefore, it seems that accentual information does not depend on the absolute values of the formant frequencies but like linguistic information it depend partly a matter of the relative formant structure of vowels. In other words, both aspects of the phonetic quality of these vowels depend partly on the relative formant structure. Secondly, there is tentative evidence that subjects belonging to different sociolinguistic groups gave different responses to some of the test material. The personal information conveyed by vowels depend partly on the absolute values of the formant frequencies.

Theory of adaption level has been put forward to explain the finding in other fields of sensory judgement (**HELSON, 1948**). The stimulus to which at any time the neutral response is to be given is called the adaption level for that time. The application of adaption level theory to vowel judgements is a series of speech sounds in which the formant position was high and raise the adoption level and so would cause a given test word to be judged as having a value of formant one lower than the adaption level. Considerable general interest attaches to the degree of success achieved by adaption level theory in explainations of vowel judgements.

In summary various aspects of vowel quality are listed as below:

(1) The acoustic quality of most vowel sounds can be specified by the frequencies of their first two or three formants.

(2) This is not true of vowel which are called close vowels, and back vowels. It is not easy to analyze these vowels in terms of their formants.

(3) The perceptual quality of a vowel also depends on the relationship between the formants of other vowels produced by the same speaker.

(4) The listener to speech uses his past experience to form an adaptation level, the immediate past experience of a particular voice being the most important factor in this process.

There are a large number of psychophysical

experiments on the perception of quality differences. So far researchers do not know which are the important auditory cues for a listener assessing the qualities of vowels and speech sounds. According to the traditional theory, there are three main parameters of vowel quality which are independently variable: the position of the highest point of the tongue in the close-open dimension, and in the front-back dimension; and degree of lip rounding. There is at the moment no other way in which vowels can be specified with equal accuracy.

Chapter 3

SPECTRAL ANALYSIS AND SPECTROGRAPHIC STUDY OF SPEECH SOUNDS

SPECTRAL ANALYSIS OF SPEECH:-

There are two advantage of frequency-domain representation of speech information. First, acoustic analysis of the vocal mechanism shows that the normal mode or natural frequency concept permits concise description of speech sounds. Second, perhaps the ear makes a crude frequency analysis at an early stage in its processing. Further, the vocal mechanism is a quasi-stationary source of sound. Its excitation and normal modes change with time. Therefore, any spectral measure applicable to the speech signal should reflect temporal features of perceptual significance as well as spectral features.

SHORT-TIME FREQUENCY ANALYSIS:-

The conventional mathematical link between f(t) and F(w) is the Fourier transform-pair

$$F(w)=\int_{-\infty}^{\infty} f(t)\, e^{-jwt}\, dt$$

$$f(t)= 1/2\pi\int_{-\infty}^{\infty} F(w)\, e^{jwt}\, dw \qquad ---(1)$$

Where, f(t)= an aperiodic time function, and

F(w)= its complex amplitude-density spectrum

For the transform to exist, $\int_{-\infty}^{\infty}|f(t)|\, dt$ must be finite.

Generally, a continuous speech signal neither satisfies the existence condition nor is known over all time. Therefore, the signal must be modified so that its transform exists for integration over known (past) values. Further, to reflect significant temporal changes, the integration should extend only over times appropriate to the quasi-steady elements of the speech signal. A running spectrum is desired,

with real-time as an independent variable, and in which the spectral computation is made on weighted past values of the signal. Such a result can be obtained by analyzing a portion of the signal seen through a specified time window, or weighting function. The window is chosen to ensure that the product of signal and window is Fourier transformable. Suppose h(t) is the weighting function such that h(t)=Ø; for t<Ø, then, from equation (1) the desired operation is

$$F(w,t) = \int_{-\infty}^{t} f(\lambda)\, h(t-\lambda) e^{-jw\lambda}\, d\lambda \text{ , or,}$$

$$F(w,t) = e^{-jwt} \int_{0}^{\infty} f(t-\lambda)\, h(\lambda) e^{jw\lambda}\, d\lambda \qquad \text{-------(2)}$$

The short-time trasform, so defined, is the convolution

$$[f(t)e^{-jwt} * h(t)], \text{ or, } e^{-jwt}[f(t) * h(t)e^{jwt}].$$

If the weighting function h(t) is considered to have the dimension of $\sec^{-1}$ (i.e., the Fourier transform of h(t) is dimensionless), then $|F(w,t)|$ is a short-time amplitude spectrum with the same dimension as the signal. Like the conventional Fourier trasform, F(w,t) is generally complex with a magnitude and phase. For example, $|F(w,t)|\, e^{-j\theta(w,t)}$, where $\theta(w,t)$ is the short-time phase spectrum.

MEASUREMENT OF SHORT-TIME SPECTRUM:-

Equation (2) can be written as

$$F(w,t) = \int_{-\infty}^{t} f(\lambda) \cos w\lambda\ h(t-\lambda)\, d\lambda - j\int_{-\infty}^{t} f(\lambda) \sin w\lambda\, h(t-\lambda)\, d\lambda$$

$$= [a(w,t) - jb(w,t)] \qquad \text{--------(3)}$$

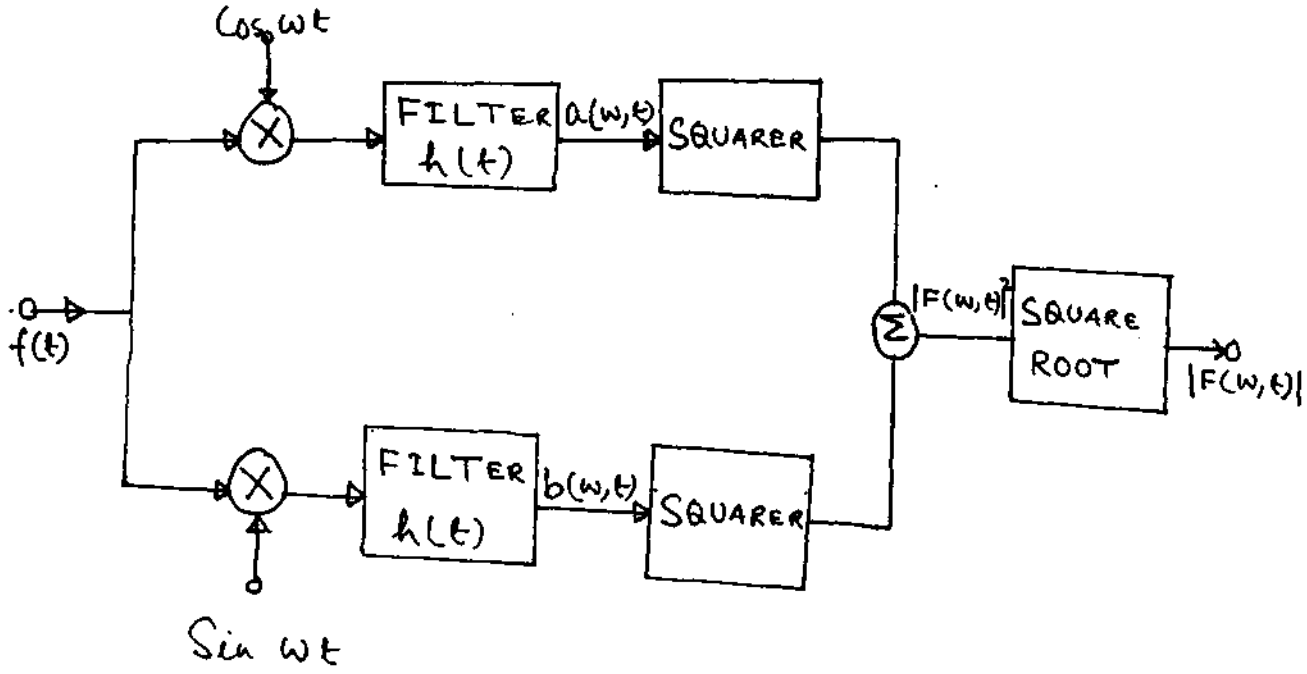
Further,

$$|F(w,t)| = [\, F(w,t)\, F^{*}(w,t)\,]^{1/2}$$

$$= (a^2 + b^2)^{1/2} \qquad \text{----------(4)}$$

and $\theta(w,t) = \tan^{-1} b/a$.

Where $F^*(w,t)$ is the complex conjugate of $F(w,t)$. $|F(w,t)|$ is a scalar, whereas $F(w,t)\,F^*(w,t)$ is formally complex and $|F(w,t)|^2$ is the short-time power spectrum. The measurement of $|F(w,t)|$ can be implemented by the operations shown below:



Multiplication by **cos wt** and **sin wt** shifts the spectrum of f(t) across the pass-band of filter h(t). Frequency components of f(t) lying close to w produce difference-frequency components inside the low - pass band and yield large outputs from the h(t) filter. Quadrature versions of the shifted signals are squared and added to give the short-time power spectrum $|F(w,t)|^2$.

Alternatively, equation (2) can be written as.

$$F(w,t) = e^{-jwt} \{\int_0^\infty f(t-\lambda)\, h(\lambda) \cos w\lambda \, d\lambda + j\int_0^\infty f(t-\lambda)\, h(\lambda) \sin w\lambda \, d\lambda \} \quad \text{--------(5)}$$

$$= [a'(w,t) + j\, b'(w,t)]e^{-jwt}$$

Practical measurement of the short-time spectrum $|F(w,t)|$ by means of a band-pass filter, a rectifier and a smoothing network is given as below:

The measurement method of above figure is precisely the one used in the sound spectrograph and in most filter-bank spectrum analyzers. It is usually the method used to develop the short-time spectrum in vocoders and in several techniques for automatic formant analysis.

AIM OF THE SPECTROGRAPHIC STUDY:-

Because of the highly complex character of the speech waves, inadequacy of the knowledge of how the intelligence is imbedded in the acoustic and other parameters and the statistical variations associated with biological process involved, this apparently simple problem has engaged the attention of researchers for almost a century.

We shall discuss acoustic-phonetic features of human speech sound. At the acoustic level, the analytical methods employed before the last world war certainly contributed to store information on the acoustic cues of

speech but these analyses in some cases led to erroneous conclusions. The great impact on speech research come with the development at the Bell laboratories in 1945 of the sound spectrograph.

Spectral analysis of speech means frequency-domain representation of speech information. Sound spectrograph provides a convenient means for permanently displaying the short-time spectrum of a sizeable duration of signal. Its choice of time windows is made to highlight important acoustic and perceptual features such as formant structure, voicing, friction. stress and pitch.

In spectrograph the sample is first recorded on a magnetic disc. The recorded signal is played repeatedly and passed through heterodynically variable band pass filter. The intensity or energy output is registered on electric sensitive facsimile paper. An electromechanical link insures that tuning of the filter and the position of writing stylus are changed on each operation. On a sound spectrogram time is represented on the horizontal axis, frequency along the vertical axis and variation in intensity as darkness. Two band pass filter widths are generally used in sound spectrogram, **(1) 45 Hz**: called as narrow band, and **(2) 300 Hz**: called as broad band. The narrow band is used to investigate frequency composition in detail, i.e., harmonics of voiced sounds and the manner of harmonic variation with time. The wide band is used to investigate broad frequency and time variation resulting from selective modulation produced by variation in the vocal cavaties during the

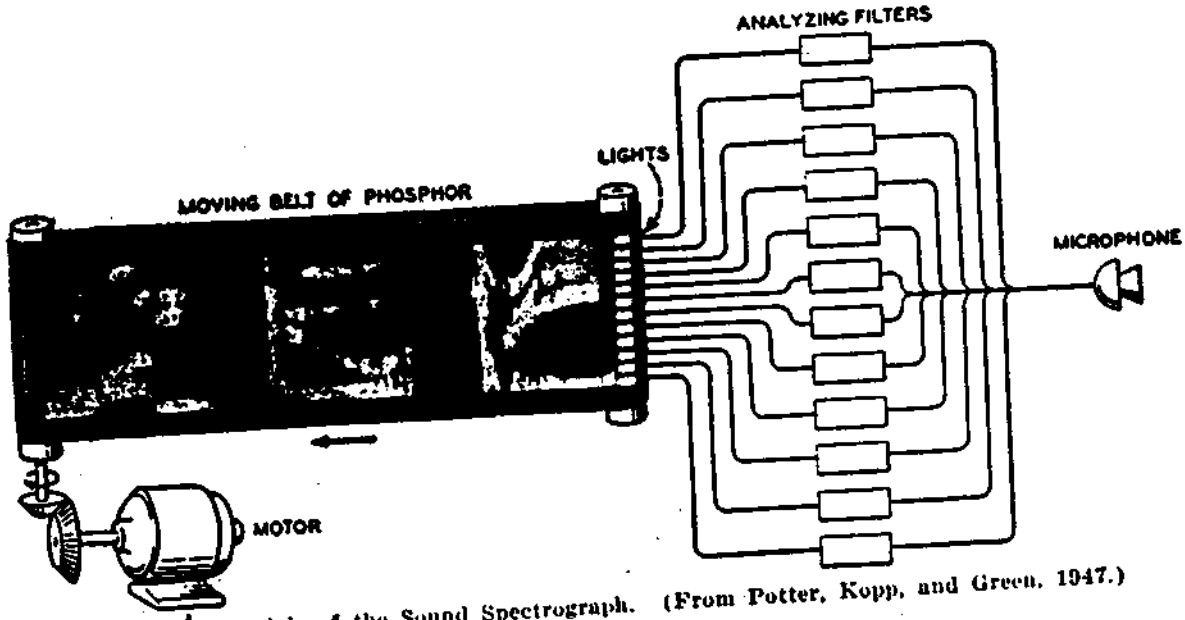


FIG. 3-1: Principle of the Sound Spectrograph. (From Potter, Kopp, and Green, 1947.)

formation of sounds. The broad band pattern is also used for pitch determination, which is accomplished by measuring intensity and frequency of vertical stirations. The spectrograph has been used to obtain the acoustic features of speech such as formant - frequency, band width, trasition and duration and other acoustic events. **POTTER, GREEN, and KOPP** made a general survey of all types of English sounds. A comprehensive survey of formant frequency and amplitudes for English vowels were made by **PETERSON and BARNEY.** Some definitions related to spectrographic measurements are listed below:

FORMANT FREQUENCUY:-

The frequency of a formant is the position on the frequency scale of the peak of the spectrum envelope drawn to enclose the peaks of the harmonics. When two formants come closure or when the formant to be measured is very low in frequency only one side of the formant peak may be visible and the estimate has to be based solely on this information. In such cases it pays to go to the broad band spectrum and determine the centre of the formant band. This method gives more accurate values than may be expected. An experienced investigator may take all his data on formant frequencies from the broad spectrogram.

In spite of its apparent lack of rigidity, this method for determining formant frequencies is preferable to the centre of gravity measurements proposed by **POTTER and STEINBERG (1952).** The application of their formula may be

give considerable errors for asymmetrical configurations of harmonics, as in the case of very low frequency formants or when two formants come close. The formant frequency is given by the following relation:

$$F = \frac{\Sigma Wi\ Fi}{\Sigma Wi}$$

Where, Fi =frequency of the ith component, and

Wi = a weighting factor = (Ai/AØ)

Ai = the amplitude of the i^{th} component, and

AØ = the amplitude of the dominant or maximum component.

This is taken as precise definition of a formant frequency.

FORMANT LEVEL:-

The formant level can be defined as the peak value of the formant envelope at the frequency of the formant. For the case of formant frequency occuring at the exact frequency of a harmonics, the envelope level equals that of the centre harmonic.

FORMANT BANDWIDTH:-

The formant bandwidth for an isolated formant that is well defined by harmonics of a low pitch is determined by the following procedure:(1) Determine those points of the spectrum envelope on both sides of the peak where the level of the envelope is 3 db below the peak level. The bandwidth is the frequency difference between the two points.(2) For the case of a high voice fundamental frequency or when two formants come too close or the formant under investigation is

too low in frequency, it may not be possible to measure the bandwidth and any estimate would tend to give too large a value. This concept of bandwidth should not be confused with the frequency width of the apparent base of the formant. Also it should not be confused with the width of the bands in the spectrogram which are essentially determined by the properties of the analyzer.

According to **FANT (1956)**, the natural range of variation of the formant frequencies for non-nasal voiced sounds uttered by an average male speaker is as follows:

F1 : 150 -850 cps

F2 : 500 - 2500 cps

F3 : 1700 - 3500 cps

F4 : 2500 - 4500 cps

For females the formants are on the average about 17% higher and for children even more higher. Typical values for FØ in speech are 120 cps for male voices, 220 cps for female voices and about 300 cps for children of about 10 years of age. The mean value of pitch within each of the male and female groups varies of course according to voice characteristic.

A vowel or any voiced sound can be regarded as the response of the vocal tract system to the voice source at the vocal cords. In other words, the vowel has the properties of the voice source modified by the filtering action of the vocal tract. A third factor enters also, and that is the radiation at the lips of the speaker. It can be thought of as a part of the filter function, or it can be handled seperately.

If a decibel scale is used, the vowel spectrum is merely the sum of the seperate frequency curves. If a linear scales are used, the vowel spectrum is the product of each of the frequency curves for the parts. The periodicity of the vowel, i.e., the harmonic structure of the spectrum, originates from the vocal cord source, the formants come from the vocal tract filter function, and the change in formant levels and the overall slope of the spectrum envelope that occur in cases when formant frequencies and bandwidths are kept constant, depend on the slope of the envelope of the voice source spectrum. These relations may be expressed by the following formula:

Sound speech = (source) . (filter function) . (radiation).

FORMANT ANALYSIS OF SPEECH:-

Formant analysis of speech can be considered a special case of spectral analysis. The objective is to determine the complex natural frequencies of the vocal mechanism as they change temporally. The changes are conditioned but the articulatory deformations of the vocal tract. One approach to such analysis is to consider how the modes are exhibited in the short-time spectrum of the signal.

Since the damping or dissipation characteristics of the vocal system are relatively constant and predictable, especially over the frequency range of a given formant. Therefore, generally more interest attaches to the temporal variations of the imaginary parts of the complex formant frequencies than to the real parts. Nevertheless, an

adequate knowledge of the real parts, or of the formant bandwidths, is important both perceptually and in spectral analysis procedures.

The system function approach to speech analysis, aims at a specification of the signal in terms of a transimission function and an excitation function. If the vocal configuration is known, the mode pattern can be computed, and the output response to a given excitation can be obtained. In automatic analysis for encoding and transmission purposes, the reverse situation generally exists. One has available only the acoustic signal and desires to analyze it in terms of the properties of the source and the modes of the system. One main difficulty is in not knowing how to seperate uniquely the source and the system.

The normal modes of the vocal system move continuously with time, but they may not always be cleary manifest in a short-time spectrum of the signal. A particular pole may be momentarily obscured or suppressed by a source zero or by a system zero arising from a side-branch element (such as the nasal cavity). The short-time spectrum generally exhibits the prominent modes, but it is often difficult to say with assurance where the low-amplitude poles or significant pole-zero pairs might lie.

Further complicating the situation is due to the fact that the output speech signal is generally not a minimum-phase function. If it were, its phase spectrum would be implied by its amplitude spectrum. The vocal-tract

transimission is minimum phase for all conditions where radiation takes place from only one point, i.e., mouth or nostril. For simultaneous radiation from these points it is not. These factors conspire to make accurate automatic formant analysis a difficult problem.

FORMANT - FREQUENCY EXTRACTION:-

The enevelope of the amplitude spectrum has a maximum at a frequency equal essentially to the imaginary part of the complex pole frequency. The formant frequency might be measured either by measuring the axis-crossing rate of the time waveform, or by measuring the frequency of the peak in the spectral envelope. If the bandwidth of the resonance is relatively small, the first moment of the amplitude spectrum,

$$\bar{f} = \frac{\int f\, A(f)\, df}{\int A(f)\, df}$$

might also be a resonable estimate of the imaginary part of the pole frequency.

The resonances of the vocal tract are multiple. The output time waveform is, therefore, a superposition of damped sinusoids and the amplitude spectrum generally exhibits multiple peaks. If the individual resonances can be suitably isolated, by appropriate filtering, the axis-crossing measures, the spectral maxima and the moments might all be useful indications of formant frequency. One such approach is the detailed fitting of an hypothesized spectral model to the real speech spectrum.

AXIS - CROSSING AS MEASURE OF FORMANT FREQUENCY:-

One of the earliest attempts at automatic tracking of formant frequencies was an average zero-crossing count (**PETERSON**). The average density of zero-crossing of the speech wave and of its time derivative was taken as approximations to the first and second formants, respectively. The reasoning was that in the unfiltered, voiced speech the first formant is the most prominent spectral component. It is therefore expected to have the strongest influence upon the axis-crossing rate. In the differentiated signal, on the other hand, the first formant is de-emphasized and the second formant is dominant. The results of these measures were found to be poor, so the method failled to give acceptable precision.

SPECTRUM SCANNING AND PEAK - PICKING METHODS:-

Another approach to real-time automatic formant tracking is simply the detection and measurement of prominences in the short-time amplitude spectrum. Atleast two methods of this type have been designed and implemented (**FLANAGAN, 1956**). One is based upon locating points of zero slope in the spectral envelope, and the other is the detection of local spectral maxima by magnitude comparision.

In the first, a short-time amplitude spectrum is first produced by a set of bandpass filters, rectifiers and integrators. The outputs of the filter channels are scanned rapidly (order of 100 times per second) by a sample-and-hold circuit. This produces a time fucntion which is a step-wise representation of the short-time spectrum at a

number of frequency values. For each scan, time fuction is differentiated and binary-scaled to produces pulses marking the maxima of the spectrum. The marking pulses are directed into seperate channels by a counter where they sample a sweep voltage produced at scanning rate. The sampled voltage are proportionals to the frequencies of the respective spectral maxima and are held during the remainder of the scan. The resulting stepwise voltages are subsequentiy smoothed by low-pass filtering.

The second method segments the short-time spectrum into frequency ranges that ideally contain a single formant. The frequency of the spectral maximum within each segment is then measured. In the simplest form the segment boundaries are fixed. Additional control circuitry can automatically adjust the boundaries so that the frequency range of a given segment is contingent upon the frequency of the next lower formant. The normalizing circuit clamps the spectral segment either in terms of its peak value or its mean value. The maxima of each segment are selected at a rapid rate (100 times per second) and a voltage proportional to the frequency of the selected channel is delivered to the output. The selections can be time-phased so that the boundary adjustments of the spectral segments are made sequentially and are set according to the measured position of the next lower formant.

A rough indication of the performance of the above technique showed that output follows F1 of vowels within

+ 150 cps greater 93% of the time and F2 within + 200 cps greater 91% of the time (**FLANAGAN, 1956**). Because of its simplicity and facility for real-time analysis this method has proved useful in several investigations of complete formant-vocoder systems (**FLANAGAN and HOUSE**).

DIGITAL COUMPUTER METHODS FOR FORMANT EXTRACTION:-

Due to enhanced ability of the computer to store and high speed manipulation, advantage is extended to all phases of speech processing. The relations between sampled-data systems and continuous systems permit simulation of complete tramission systems within the digital computer.

The digital analyses which have been made for speech formants have been primarily in terms of operations on the spectrum. The spectrum either is sampled and read into the computer from an external filter bank, or is computed from a sampled and quantized version of the speech waveform. One analysis procedure locates the first and second formants in voiced segments. The formant tracking procedure is basically a peak-picking scheme. However, many programmed constraints are included to exploit vocal tract characteristics and limitations.

The automatic analyzing procedure, prescribed by a program, first locates the absolute maximum of each spectrum. A single formant resonances is then fitted to the peak. The single resonance is pqsitioned at a frequency corresponding to the first moment of that spectral portion lying, say, from zero to 60 db down from the peak on both sides. The single formant resonance is then inverse-filtered

from the real speech spectrum by substracting the log-amplitude spectral curves. The operation is repeated on the remainder until the required number of formants are located. Since the peak-picking is always accomplished on the whole spectrum, and accurate results can be obtained on running speech. The analysis is done in real time, and the computer can be used as the formant-tracking element of a complete formant-vocoder system (**COKER and CUMMISKEY**).

The formant frequencies correspond closely with the resonance peaks in the smoothed spectrum. Therefore, a good estimate of the formant frequencies is obtained by determining which peaks in the smoothed spectrum are vocal tract resonances. Constraints on formant frequencies and amplitudes, derived from a three-pole model of voiced sounds, are incorporated into an alogrithm (**FUJISAKI**) which locates the first three formant peaks in the smoothed spectrum.

MEASUREMENT OF FORMANT BANDWIDTH:-

The bandwidth of the formant resonances-or the real parts of the complex poles - are indicative of the losses associated with the vocal system. Not only are quantitative data on formant bandwidths valuable in corroborating vocal tract calculations, but a knowledge of the damping is important in the proper synthesis of speech.

A number of measurements have been made of vocal tract damping and formant bandwidth. The measurements divided mainly between two techniques; either a measure of a resonance width in the frequency domain, or a measure of a

damping constant (or decrement) on a suitably filtered version of the speech time waveform. The damping constant, σ, for the wave and its half-power bandwidth, Δf, are related simply as

$$\sigma = \pi \Delta f = \frac{\ln A_2/A_1}{(t_2 - t_1)}$$

Where,

σ = damping constant,

Δf = half power bandwidth,

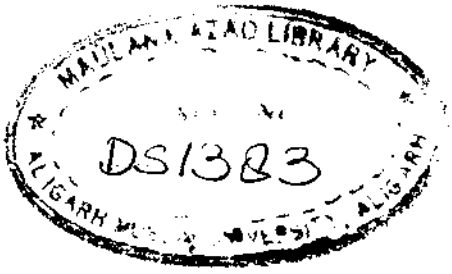
A_1 = amplitude of formant at time t_1,

A_2 = amplitude of formant at time t_2.

Formant bandwidths can also be effectively measured from a frequency response of the actual vocal-tract (**FUJIMURA**). A sine wave of volume velocity is introduced into the vocal-track at the glottal end by means of a throat vibrator. The pressure output at the mouth is measured as the input source as change in frequency.

Chapter 4

REVIEW AND EXPERIMENTAL STUDY



REVIEW:-

The study of different aspects of vowel sounds in different languages is updated continuously. To determine articulatory data from measurements on the speech wave two stages are required: (1) an acoustic analysis of the speech signal must be made, and (2) the prabable articulatory configuration that yields this acoustic must be computed **(PRTERSON, 1952). STEVEN, K. N. and HOUSE, A. S. (1)** advanced these data and developed a set of parameters (rØ - constriction size, dØ - distance from the glottis to the point of constriction, A/l - constriction of cross - sectional area) that yield a simple yet reasonably accurate specification of the articulation of vowel sounds. They also establishes relations between the parameters that described the articulatory events and the formant frequencies that describe the acoustic events. The effects on the vowels of the variation of the three parameters (rØ, dØ, A/l) is interpreted as adding to the face validity of this simplified description. They concluded that, the proposed scheme produces configuration of the vocal tract analog that in turn yield formant frequencies appropriate to the end product desired, and that the relationship among the vowels so produced are similar to those regarded by students of speech as characteristic of human speech.

POLS, L. C. W.; KAMP and PLOM (2) worked on perceptual and physical space of vowel sounds in which experiments were carried out to investigate the correlation between the perceptual and physical space of 11 vowel sounds.

The signals were single periods out of the constant vowel part of normally spoken words of the type h(vowel)t, generated continuously by computer. Pitch, loudness, onset, and duration were equalized. These signals were presented to 15 subjects in a triadic - comparison procedure, resulting in a commulative similarity matrix. Multidimensional scaling of this matrix resulted in a three dimensional perceptual space with 1.6 stress. The signals were also analyzed physically with 1/3 - octave band filters. Principal - components analysis of the decibel values per frequency band indicated that three dimensions accounted for 81.7% of the total variance. Matching the perceptual and the physical configurations to maximal congruence yielded an excellent result with correlation coefficients of Ø.992, Ø.971 and Ø.742 along the corresponding dimensions. The formant frequencies and levels were correlated also with both configurations.

Further, **FANT (3)** has advanced a notation of non-uniform scaling in which the vowels of males and females are uniformaly scalable within their genders, but not to each other. Because the effect is small and still a subject of discussion **(NORDSTROM and LINDBLOM).**

HIRSH (4) looked more closely at the spectrum of a signal by narrowing the bandwidths of the spectrum analyzer, the less he can observe in the time domain, and argued that: "Time is the dimension on within which patterns are articulated for hearing in a manner analogous to the way in

which space is patterning dimension for vision".

KLEIN, W.; PLOMP and POLS (5) were analyzed twelve Dutch vowels, each pronounced by 5Ø male speakers in 18 - filter bands comparable in bandwidth with the ear's critical band. The result showed that the confusion between the vowel types were basis for a multidimensional scaling **(KRUSKAL)** to construct a perceptual configuration of the vowels.

SCHOUTEN et al (6) worked on "residue pitch" and his result indicated that temporal analysis may be important for the perception of fundamental frequency.

RUPF, J. A., HUGES and HOUSE (7) studied about the effect of switching on the recognition of speech sounds in which signal was switches from one ear to other and demonstrated that the place at which switching occured had essentially no effect on the identification of either the consonants or the vowels in the stimuli.

SCOTT, B. L. (8) studied about the Temporal Factors in Vowel Perception. During his study the perceptual role of the temporal fine structure of vowel waveforms was investigated in fine experiments. The results indicated that the perceptual system was responding to either a change in number of cycles of F1 per fundamental period or a change in harmonic structure of the seconds. The hypothesized temporal cue was then used in systhesizing vowel-like sounds which, while not differing in formant centre-frequency or harmonic structure, did differ in temporal structure.

Working on perceing vowels in Isolation and in Consonant context **(9) DIEHL, R. L., MCCUSKER, S. B., and**

CHAPMAN, L. S. showed that vowels tend to be identified more accurately in consonantal context than in isolation. Thus consonantal---bounded vowels would appear to be acoustically less distintive than isolation vowels because one posibility is that isolation vowels may be produced with less durational varition across categories than consonantal - bounded vowels. Similar results has been given by **STRANGE et al,(1976); BISCHEFF (1976); STRANGE et al (1979); GOTTFRIED & STRANGE et al (198Ø).** The opposite outcome might have been predicted since formant trajectories of consonant - bounded vowels often to reach the frequencies characteristic of vowels produced in isolation (**LINDBLOM,1963; STEVENS & HOUSE, (1963).**

The role of intrisic fastors determining perceived degree of vowel openess was determined by **TRAUNMIILLER, H. (1Ø).** In order to determine the role of F1 and FØ, one formant vowels, covering a wide range of fundamental and formant frequencies, were identified by 23 subjects who were native speakers of a Bavarian dialect in which five degree of openess occur distinctively. It was found that the distance between widely spaced formants, as between F2 and F1 in front vowels, is not crucial for vowel identification. Most of the distinctive features of speech sounds apparently constitute dimensions along which only a binary distinction is used. Along the dimension of "openess" in vowels, more than two distinction occur in many language. In this investigations an attempt is made to examine how this

dimension is percieved. The investigation is limited to the role of intrisic factors. The aim of this experiments is to establish more reliable decisive criteria for the identification of one - formant vowels. The experiment was plan to cover the whole range of variations in FØ (5Ø to 7ØØ Hz) and F1 (15Ø to 148Ø Hz) that can be observed in natural speech.

BERNSTEIN, J. (11) did the experiments in which he try to find out that, can the acoustic properties of vowels, in particular, formant frequencies, and formant amplitudes, be scaled so that the auditory relations among vowels are representable by their positions in formant frequuency space? In experiments, carefully selected sets of five formant vowel like sounds were presented to listeners. Test results demonstrate that if inter-vowel distances are organised on a formant-by- formant basis,no scaling of formant frequencies can possibly reproduce the correct ordinal properties of listeners data.

PORT, D. K. (12) examined the acoustic correlates of place of articulation in the voiced formant transitions from natural speech during his study. The results indicated that the information contained in the formant transitions in these natural stop-vowel syllables was not sufficient to distinguish place across all the vowel contexts studied. Most of the research supporting this point of view has come from experiments using synthetic speech conducted at Haskins laboratories **(COOPER et al.,1958, LIBERMAN et al.1954;HARRIS et al.,1958)**. The outcome of these perceptual studies

indicated that place of articulation could be specified from formant transition information alone, although the acoustic correlates of this information were not invariant over vowel context.

An adequate theory of vowel perception must account for perceptual constancy over variations in the acoustic structure of coarticulated vowels contributed by speaker, speaking rate, and consonantal context was tested by **STRANGE, W., JENKINS, J. J.,& JOHNSON, T. L.(13).**Results of identification tests by untrained listeners indicated that dynamic spectral information, contained in initial and final transitions taken together, was sufficient for accurate identification of vowels even when vowel nuclei were attenuated to silence. The research reported here addresses the problem of the correspondence between the acoustic signal and phonetic percept for a major class of English phoneme, the vowels. Vowels have traditionally been differentiated in articulatory terms by the static vocal tract shapes attained by positioning the tongue, jaw, and lips in different configurations.

The identifiability of isolated vowels (/V/) was compared to that of vowels in consonantal context (/pVp/) when subjects performed a monitoring task **(14)** by **RAKERD, B., VERBRUGGE, R. R., & SHANKWEILER, D. P.** On average, resulting errors occured significantly less often in the /pVp/ condition, consistent with the previous finding (**STRANGE et al., 1976; MACCHI, 1980; DIEHL et al., 1981; ASSMANN et al.,**

1982) that vowel perception may be aided by consonantal context. This beneficial effect of context was found to be restricted to the class of open vowels with perception of the close vowels being somewhat hindered by context. The error data for misses also showed an interaction between context and vowel height.

Under many circumtances, listners identify vowels in consonantal context more accurately than vowel in isolation. The purpose of the **GOTTFRIED, T. L. & CHEW, S. L.** (15) experiment was to examine the effect of FØ and consonant context on the intelligibility of vowels sung by a male singer in two different vocal registers. The countertenor voice was chosen because of the opportunity to examine in one singer (and one vocal tract), two different voices whose ranges overlap. Each vocal register might have distinct effects on the intelligibility of vowels. The inteligibility of the gated vowels should be considerably lower than that of the CVC syllables.

Acoustic analyses are reported which show that the spectral properties of stuttered vowels are similar to the fluent vowel, so it would appear that the stuttered are articulating the vowel appropriately. The stuttered vowels are low in amplitude and short in duration. In two experiments of **HOWELL, P. & VAUSE, L.** (16) the effects of amplitude and duration on perception of these vowels are examined. These experiments lead to the conclusion that low amplitude and short duration are the factors that cause stuttered vowels to sound like / schwa/. The formant

frequencies of stutters for fluently and dysfluently produced vowels are compared. The analysis show that the formants of the stuttered vowels are located at about the same frequencies as the fluent vowels and thus the vowels are articulated correctly.

The purpose of the Piecewise plane vowel formant distributions across speakers by **BROAD, D. J. (17)** was to examine the forms of the clustering of the first three vowel formant frequencies for a number of speakers and to test the hypothesis that each clustering can be represented by equation (1)

$$\alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3 + \alpha_4 / k = 0,$$
$$\text{for } F_2 > a_1 F_1 + a_2 / k, \qquad \text{——— (1)}$$
$$\beta_1 F_1 + \beta_2 F_2 + \beta_3 F_3 + \beta_4 / k = 0,$$

Where the α's, β's and a's are data determine constants. The hypothesis of uniform formant scaling was thus be tested by examining the clustering of formants for various speakers vowel spaces. If the clustering can be represented by equation (1), then uniform scaling is partially confirmed. If not, it is rejected.

PETERSON, G. E. & LEHISTE, I. (18) study about Duration of syllable Nuclei in English deals with the influence of preceding and following constant on the duration of stressed vowels and diphthongs in American English. A set of 1263 CNC words, pronounced in an identical frame by the same speaker, was analyzed spectrographically, and the influences of various classes of consonants on the

duration of the nucleus were determined. The residual duration differences are analyzed as intrinsic durational characteristics, associated with each syllable nucleus. The theory is tested with a set of 30 minimal pairs of CNC words, uttered by five different speakers.

HUGGINS, A. W. F. (19) studied about the jnd differences for segment duration in natural speech. He showed that subjects are much more sensitive to changes in vowel duration than to change in consonant duration. He also studied about the perception of temporal phenomena in speech and suggested that the perception of timing in natural speech is based on events at the syllabic level rather than at the segment level.

Effect of Structure and Duration of Vowels on fricative voicing was studied by **SOLI, S. D. (20)**. The duration of vowels specify several types of linguistics information. Thus multiple duration cues may be decoded from the same temporal interval of the signal, a situation which whould create perceptual ambiguity. Four perceptual studies of postvocalic fricative voicing examined how information in the duration and dynamic structure of the vowel might combine to resolve these ambiguities. This research focuses on the latter of the two preceding alternatives, examining the hypothesis that as the duration of a vowel is modified according to a linguistic timing rule its dynamic spectral shape varies in a regular fashion. The experimental hypothesis stated concisely is that the dynamic structure of the vowel provides cues which combine with duration cues to

specify unambiguous linguistic information. According to the present hypothesis, it is in this fashion that linguitic information may be encoded in both the dynamic spectral structure and duration of the vowel.

BLADON, R. A. W. and LINDBLOM, B. (21) worked on Modeling the Judgement of Quality Differece. The hypothesis of this study is that the auditory cues relevant to listeners judgement of vowel quality are a spectral representation of loudness density versus pitch. A model is described that generates such patterns for steady - state vowels. This model is combined with a measure of auditory perceptual distance which, operating on pairs of vowels, treats each stimulus representation as a single spectral shape.

A commonly held assumption about memory for speech is that auditory memory is referred to only if phonetic memory does not contain the information needed for a particular trail. **COWAN, N. (22)** study provides addition data to help to determine how auditory and phonetic memory are used in a vowel discriminition task, and what happens during memory decay. For this task two experiments was conducted. Experiment one was conducted to determine whether performance levels decline at similar rates on between - and within - category AX vowel comparision trials when certain methodological problems are removed. Experiment two demonstrated that in the AX task there is a vowel order effect, but that this (vowel order effect) increased across interstimulus decay intervals, in contrast to their findings.

The result can be accomodated with a model in which the memory for a vowel is represented as a small, bounded area within the vowel space, an in which memory decay is represented by the expansion of that bounded area over time.

PETERSON, G. E. & BARNEY, H. L. (23) discussed some of the control methods that have been used in a study of the vowels. In the plan of the study a list of words was presented to the speaker and his utterance of the words were recorded with a magnetic tape recorder. The list contained ten monosyllabic words each begininning with [h] and ending with [d] and differing only in the vowel. The words used were heed, hid, head, had, hod, hawed, hood, who'd, hud, and heard. The order of the words was randomized in each list, and each speaker was asked to pronounce two different lists. The purpose of randomizing the words in the list was to avoid practice effects which would be associated with án unvarying order. A total of 76 speakers, including 33 men, 28 women, and 15 children, each recorded two lists of 1Ø words, making a total of 152Ø recorded words. Majority of speakers belongs to different parts of united states and they spoke General American.

A given lists of words, recorded by speakers, were randomized and were presented to a group of 7Ø listeners in series of eight sessions and each listener was asked to write down what he heard on a second list (list II). 32 of the 76 speakers were among the 7Ø observers. A comparision of list I and list II would reveal occasional differences, or disagreements, between speakers and listeners.

The acoustical measurement were made with the sound spectrograph; to minimize measurement errors, a method was used for rapid callibration of the recording and analyzing apparutus by means of a complex test tone. For example, list I might be played into an acoustic measuring device and the outputs classified according to the measured properties of the sounds into a list III and the list I, II & III will differ in some words depending upon the characteristics of the speaker, the listener, and the measuring device. Statistical techniques were applied to the results of measurements, both of the callibrating signals and of the vowel sounds. On similar pattern results of acoustical mesurement have been reported by **FANT (1956)** for Swedish vowels and **MAJUMDER, D . D., DUTTA, A. K. & GANGULI, N. R. (1973)** for Hindi vowels.

In an experiment by **FLANAGAN and SASLOW (24)** it appeard that even when no noise is present, an effect of the vowel spectrum can be found. They assumed that a shorp low - frequency formant would mask more of the harmonics that convey information on the change in FØ than would a shallower formant.

In a study on the effect of consonantal context on formant frequencies of vowels, **STEVENS and HOUSE (25)** have found that the extent to which various consonantal contexts influence the formant frequencies differs considerably from one vowel to the other. And consonantal context gives systematic shifts in formant frequencies. The shift depends

on place of articulation, manner of articulation, and voicing characteristics of adjacent consonants.

Traditionally, the formant frequencies are regarded as the most important characteristics of the frequency spectra of vowel. **PLOMP, POLS and GEER (26)** discussed about the Dimensional Analysis of vowel spectra and reached the conclusion that it is possible to approach the differences between vowel spectra in a more general way by means of a dimensional analysis. For a particular vowel, the sound pressure levels in each of a number of frequency pass bands can be considered as coordinates of a point in a multidimensional euclidean space. Different vowel spectra will result in different points. Frequency spectra of 15 Dutch vowels were determined with 18 bandpass filters (10 speakers). The analysis indicated that the "cloud" of 150 points can be described by four independent dimensions that are linear combinations of the original 18. The percentage of total variance explained by these dimensions were 37.2%, 31.2%, 9.0% and 6.7%, respectively. This approach presents interesting perspectives for the develoment of vowel - discrimination equipment.

POLS, L. C. W.; TROMP, H. R. C. and PLOMP, R. (27) studied about the frequency analysis of Dutch vowels from 50 male speakers, the vowels were spoken by 50 male speakers in a h(vowel)t context and the frequencies and the levels of the first three formants of 12 Dutch vowels were measured statistically. Statistical analysis of these formant variables confirmed that F1 and F2 are the most appropriate

two distinctive parameters for describing the spectral differences among the vowel sounds. Maximum likehood regions were computed and used to classify the vowels, and score of 71.3% correct classification in the log F1 - log F2 plane was obtained.

Detailed study of the spectral and durational acoustic parameters for stops in CV syllables was reported by **FANT (28)**. His data consisted of spectrograms of six initial stops combined with the nine long vowels of Swedish spoken once by a single talker. From these measurement, **FANT** arrived at two major conclusions. The first,in agreement with **DHMAN,s (1966)** study, was that a consonant locus for place could not be obtained from the spectrographic measurements. The second, which was at variance with the claims made in the Haskins perceptual studies, was that F2 and F3 formants trasition patterns did not always uniquely specify place of articulation. By measuring transitions from one subjects to, it is possible to test **FANT** hypothesis that formant transitions are not distinctive correlates of place.

BENNETT, S. (29) did experiments on vowel formants frequency characteristics of preadolescent males and females. This experiments describes the vowel formant frequency characteristics (F1-F4 of five vowels produced in fixed phonetic context) of 42 seven and eight year old boys and girls and the relationship of vocal tract resonances to several indices of body size. Results showed that the vowel resonances of male children were consistently lower than

those of females, and that the extent of the sexual differences varied as a function of formant number and vowel category.

An experiment was carried out by **SCHEFFERS, M. T. M. (3Ø)** investigating the relationship between the just noticeable difference of fundamental frequency (JND & FØ) of three stationary synthesized vowel sounds in noise. The S/N ratios were measured at which listeners could just discriminate a series of changes in FØ in the range from 1Ø% to Ø.5%. Similar measurements were obtained for pulse trains. Using this measure it was found that a given change in the fundamental frequency of a pulse train could be discriminated at a lower S/N ratio than in a pulse tone with a frequency equal to that fundamental. The results for the vowel sounds were found to be in between those for a low frequency pulse tone and those for a pulse train. This experiment was carried out as a part of a research project investigating the role of pitch in the perceptual seperation of speech sounds from an interfering background. The pitch of a signal harmonic was the cue most often used.

RECORDING AND LISTENING PROCEDURES:-

A list of words of 1Ø most frequent vowels of Hindi speech is made. The list contained ten monosyllabic words each begining with [h] and ending with [d] and differing only in the vowel. The words used were / hʌḍ /, / haḍ /, / hIḍ /, / hiḍ /, / hUḍ /,/ huḍ /, / hɛḍ /, / heḍ /, / hOḍ /, and / hɔḍ /. The order of the words was randomized in each list. The purpose of randomizing the

words in the list was to avoid practice effects which would be associated with an unvarying order. Each speaker was trained to read the lists carefully in natural way with a time gap of about 3 second by the author. After that, the speaker was asked to read the two randomized lists of 10 words each these and were recorded on a magnetic tape recorder. At the time of recording care has been taken that the distance between microphone and speaker is 30 cm. approximately. The lists were recorded in free field of a partially acoustic treated room.

A total of 21 speakers, consisting 15 men, 6 women, had recorded two lists of 10 words each, making a total of 420 recorded words. All the speakers were born in Hindi speaking area in **INDIA** (Uttar pradesh and Bihar) and in the age group of 20 - 30 years. Also, all were well qualified, for example, Graduates, Postgraduates, or research scholars and had studied in Hindi medium upto atleast Highschool. Some of speakers have also good background of urdu.

The randomized words were presented to a group of 15 untrained listeners in series of three sessions. The listing group contained men only and represented much of the same dialectal distribution as did the group of speakers and also in the same age groups and same qualifications as that of the speakers.

The general purpose of these tests was to obtain an aural classification of each vowel to supplement the

speaker's classification. In presenting the words to the listeners, the procedure was to reproduce at each of three sessions, 420 words recorded by 10 speakers. Each session contained 5 adult men only. The sound level at the observer's positions was approximately 70 db and varied over a range of about 3 db at the different positions.

Each listener was given a articulation test record sheet. The listener was asked to write the one words in each line that he heard. At the time of writing, no choice of leaving blank was given to the listeners, instead, the listeners were forced to write a vowel in Hindi orthography in each line what he heard and judged correspondingly. A comparision of both, randomized lists recorded by speakers and in articulation test records sheet written by listeners would have occasional differences, or disagreements, between speakers and listeners. The seating positions of the listeners were randomized so that it would cancel the effect of position in the laboratory may had on the identification of the sounds, if any. These lists were played in free field of a partially acoustic treated room. The confusion matrix and the percentage of missidentification according to the place of articulation and openness of the vocal tract are shown in a **Table - 1 & Table - 2(a)** respectively. Comparision of intellegibility with **PETERSON and BARNEY (23)** are shown in **Table - 2(b).**

MEASUREMENT OF THE SPECTROGRAM:-

Spectrograms are taken on Sona - Graph 7029 - A which is available in laboratory.

TABLE — 1: Confusion Matrix of Vowel

VOWELS INTENDED BY SPEAKERS ↓	VOWELS CLASSIFIED BY LISTENERS →									
	ʌ (अ)	a (आ)	I (इ)	i (ई)	U (उ)	u (ऊ)	ɛ (ए)	e (ऐ)	O (ओ)	ɔ (औ)
ʌ (अ)	589	22	-	3	-	3	-	7	-	6
a (आ)	20	576	-	-	-	-	-	3	-	31
I (इ)	-	-	535	37	3	-	55	-	-	-
i (ई)	-	-	34	587	-	-	9	-	-	-
U (उ)	-	-	3	-	542	59	-	-	26	-
u (ऊ)	8	4	-	-	87	508	-	-	23	-
ɛ (ए)	-	-	40	16	-	-	532	42	-	-
e (ऐ)	2	-	7	3	-	-	27	587	3	1
O (ओ)	1	-	-	-	27	26	1	-	530	45
ɔ (औ)	4	39	-	-	-	-	-	-	32	555

TABLE - 2(a): For Percentage Misidentification of Vowels according to place of Articulation and Openness of the vocal tract:

Place of Articulation	% of Misidetification	% of Misidentification(*) (j - j) context	(w - w) context
Front	1Ø.62	19.3	11.3
Back	13.58	16.2	7.2

Openness of the Vocal tract	% of Misidentification	% of Misidentification(*) (j - j) context	(w - w) context
Low	7.6	17.5	6.Ø
Medium	12.6	21.2	13.7
High	13.85	1.7	Ø.Ø

(*): GUPTA, J. P. and AHMED, R (1971)

TABLE - 2(b): For comparision of % of Intellegibility with **PETERSON and BARNEY (23)** Confussion table:

Vowels	% of Intellegibility	% of Intellegibility (23)
ʌ	93.4	92.17
a	91.4	86.92
I	84.9	92.88
i	93.1	99.87
U	86.Ø	99.18
u	8Ø.6	96.53
ɛ	84.4	87.68
ɔ	88.Ø	92.74

SONA - GRAPH 7029 - A is an audio - frequency spectrum analyzer that produces permanent, graphic recordings of any type of complex wave in the range of 5 Hz to 16000 Hz. Unlike convensional spectrum analyzers, the 7029-A permits two different analyses to be displayed; the operator can select the display that most accurately shows the parameters he intends to study. The first kind display gives an over all, three dimensional picture of the signal being analyzed; frequency, amplitude and time are represented simultaneously on one display. The second type of analysis, permits the individual intensity of each frequency components to be displayed at any preselected point in time. This type of pattern is referred as a section.

This unit diplays any portion of audio in the 5 to 16000 Hz range. The input signal is first recorded on a cylinderical magnetic drum, and the played back at a high speed during the analysis process. A frequency- heterodyne technique is used for the scanning system, and there are two plug- in filters available for increased flexibility. The narrow filter emphasizes frequency resolution, and the wide filter emphasizes time resolution. A built- in calibration tone generator can provide frequency markers every 50, 500, or 1000 Hz along the frequency scale of the pattern, simply by the depressing a switch.

An adjustable AGC control is present, and can be used to extend the dynamic range of the pattern; also the darkness of the pattern can be adjusted to obtain the best

contrast. For monitoring purposes, a VU meter and a speaker can be used simultaneously either when recording the input signal or when performing the analysis.

The Sona-Graph 7Ø29-A is a completly solid state sound spectograph with all its necessary systems enclosed in one cabinet. This instrument consists of three basic systems; the drive mechanism for the turntable and drum, the electronic circuitry of the amplifier- analyser, and the regulated power supply.

ACOUSTICAL MEASUREMENTS USING A SOUND SPECTROGRAPH:-

Conventionally, manual methods are used for measuring acoustic parameters such as fundamental and formant frequencies displayed on the spectrogram. For example, the frequency is measured by dividing the scale in a linear or logarithmic manner depending on the type of analysis. The problems encountered in such measurements are, lack of absolute linearity of the frequency scale, unexpected variation in frequency due to the deposition of soot or dust on the cermet resister, changing width of the base line and other human errors.

In order to encounter these problems and for fast measurements an electronic frequency counter can be used. Such a counter has been designed and tested by **ANSARI, A. M. & AGRAWAL, S. S. at CEERI (New Delhi; 1983)** and is being used as plug-in-unit of their spectrograph. It uses carrier oscillator frequency as the reference which is a reference for marking also.

Initially, the spectrograms of the desired speech

sound is taken. The frequency range setting of the counter is set identical to the range setting of the spectrogram. Then without removing the paper from the drum, the stylus is moved to the point whose frequency is to be measured. The display on the counter now directly indicates the desired frequency. It provides a digital read - out of frequency and can be used either during or after the analysis. These measurements are independent of the frequency scale, unexpected variations in the frequency, width and position of the base line, band width of the analyzing filter, and settings of the lower and upper frequency limits of the scale magnifier unit. It has been found to work satisfactorily and the measurement time has been reduced as compared to manual methods in addition it gives accurate results.

Comparision of direct manual measurement from section and measurements using frequency counter from section and also comparision of direct manual measurement from section and measurements using formula for formant estimation $F = \frac{\Sigma Wi\ Fi}{\Sigma Wi}$ from section given by **POTTER and STEINBERG (1950)** are given in **Table - 3(a) and Table - 3(b)** respectively. From **Table - 3(a) and Table - 3(b)**, it is observed that there is a satisfactory agreement in formant frequencies of all the three methods of measurement of formant frequencies.

TABLE - 3(a): Comparision of direct manual measurements from section and measurements using frequency counter from section of some spectrograms:

Manual Measurements (Using Linear Scale) Hz (Direct from section)	Measurements (Using Frequency Counter) Hz (From section)
152	15Ø
166	165
172	167
35Ø	3Ø4
72Ø	75Ø
77Ø	78Ø
98Ø	975
1225	126Ø
147Ø	148Ø
189Ø	1938
245Ø	262Ø
259Ø	283Ø
269Ø	295Ø
294Ø	32Ø1
359Ø	3735

TABLE - 3(b): The direct manual measured formants frequencies from section compared with a formula given by **POTTER and STEINBERG (1950)** for estimated value of formant frequencies from section:

Manually Measured Formant Frequencies in Hz (Direct from section)	Formant Frequencies Measured Using Formula $F = \frac{\Sigma Fi\ Wi}{\Sigma Wi}$ in Hz (From section)
350	349
560	538
720	740
750	808
770	746
980	891
1225	1208
1470	1458
1890	1887
2345	2365
2450	2373
2590	2577
2690	2755
2940	2891
3360	3191
3590	3593
4060	4071
4200	4000

RESULTS AND DISCUSSION:

The fundamental and formant frequencies of 10 Hindi vowels for 6 males and 3 females are presented in **Table - 4 and Table - 5 respectively.** For males, the range of individual fundamental frequency is from 119 c/s for vowel ʌ (अ) and a (आ) to 200 c/s for vowel u (ऊ) and the range of average fundamental frequency is from 152 c/s for vowel a (आ) to 174 c/s for vowel u (ऊ) . For females, the individual fundamental frequency range is from 175 c/s for vowel U (ऽ) to 283 c/s for vowel u (ऊ) and average fundamental frequency range is from 198 c/s for vowel O (ओ) to 252 c/s for vowel u (ऊ). It can be observed that the spread for individual first formant frequency is from 322 c/s for vowel i (ई) to 910 c/s for vowel e (ऐ) and the range measured for average first formant frequency is from 493 c/s for vowel U (ऽ) to 715 c/s for vowel ʌ (अ) for males. For females the range for individual first formant frequency is from 409 c/s for vowel ɔ(ऑ) to 1470 c/s for vowel a (आ) and the range of average first formant is from 477 c/s for vowel O (ओ) to 1138 c/s for vowel a (आ). For individual second formant, the spread is from 993 c/s for vowel u (ऊ) to 2660 c/s for vowel e (ऐ) for males and from 1270 c/s for vowel ʌ (अ) to 2434 c/s for vowel e (ऐ) for females. For males, the average second formant frequency varies from 1230 c/s for vowel U (ऽ) to 2107 c/s for vowel e (ऐ) whereas for females, the variation is found to be from 1537 c/s for vowel ɔ (ऑ) to 2372 c/s for vowel e (ऐ). For individual third formant frequency,

TABLE — 4: Formant frequencies of Hindi Vowels (Male only)

Formant frequency	No. of Observation		ʌ (अ)	a (आ)	I (इ)	i (ई)	U (उ)	u (ऊ)	ɛ (ए)	e (ऐ)	o (ओ)	ɔ (औ)
F0 C/S	6	Minimum Fundamental frequency	119	119	133	133	162	126	123	150	132	162
		Maximum Fundamental Frequency	172	185	170	187	174	200	170	171	160	182
		Average Fundamental Frequency	154	152	160	166	169	174	153	168	148	172
		Standard Deviation	17.70	20.16	12.20	16.35	3.82	23.36	15.0	8.46	9.77	6.5
F1 C/S	6	Minimum Formant Frequency	634	425	304	322	350	361	440	490	422	500
		Maximum Formant Frequency	780	750	770	840	610	650	770	910	648	840
		Average Formant Frequency	715	606	498	569	493	504	624	650	535	680
		Standard Deviation	51.58	111.5	178.97	184.35	79.51	112.28	144.74	151.77	75.35	135.4
F2 C/S	6	Minimum Formant Frequency	1255	1125	1115	1223	1030	993	1054	1532	1050	1131
		Maximum Formant Frequency	2110	1890	1938	2355	1470	1470	1890	2660	1680	1890
		Average Formant Frequency	1578	1390	1440	1921	1230	1251	1400	2107	1270	1369
		Standard Deviation	356.81	245.97	271.12	475.65	142.2	155.05	271.72	366.93	220.60	250.4
F3 C/S	6	Minimum Formant Frequency	1750	2015	2204	3157	1680	1820	1647	2030	1680	1960
		Maximum Formant Frequency	2620	2830	3150	3850	2270	2940	3290	4170	3080	3220
		Average Formant Frequency	2298	2417	2794	3447	1966	2321	2382	3157	2388	2394
		Standard Deviation	392.61	295.12	355.30	278.78	201.85	419.3	617.17	557.09	449.24	467.1
F4 C/S	6	Minimum Formant Frequency	2250	2940	-	3780	-	-	2340	3502	2240	2814
		Maximum Formant Frequency	3201	4165	-	5040	-	-	4060	5530	4060	4200
		Average Formant Frequency	2867	3520	3290	4340	3710	3780	3328	4865	2987	3731
		Standard Deviation	436.78	459.49	-	523.83	-	-	754.37	665.0	778.09	539.3

TABLE – 5: Formant frequencies of Hindi Vowels (Female only)

rmant equency	No. of Obser-vation		ʌ (अ)	a (आ)	I (इ)	i (ई)	U (उ)	u (ऊ)	ɛ (ए)	e (ऐ)	O (ओ)	ɔ (औ)
F0 C/S	3	Minimum Fundamental Frequency	199	205	185	190	175	220	185	205	184	200
		Maximum Fundamental Frequency	201	215	257	238	250	283	232	280	210	212
		Average Fundamental Frequency	200	211	227	221	223	252	209	244	198	205
		Standard Deviation	0.81	4.18	30.72	22.17	34.23	25.72	19.20	30.72	10.78	5.09
F1 C/S	3	Minimum Formant Frequency	730	910	552	526	558	472	490	490	809	409
		Maximum Formant Frequency	910	1470	650	630	630	630	588	810	930	840
		Average Formant Frequency	827	1138	587	572	583	561	523	620	477	556
		Standard Deviation	74.08	240.0	44.43	43.29	33.47	65.93	46.19	137.35	124.33	200.62
F2 C/S	3	Minimum Formant Frequency	1270	1950	1680	-	-	-	1540	2310	1610	1420
		Maximum Formant Frequency	1820	2240	1962	-	-	-	1612	2434	1720	1750
		Average Formant Frequency	1543	2097	1821	-	1950	1680	1587	2372	1662	1537
		Standard Deviation	224.54	118.41	141.0	-	-	-	33.47	62.0	45.10	150.55
F3 C/S	3	Minimum Formant Frequency	-	-	-	-	-	-	-	-	-	-
		Maximum Formant Frequency	-	-	-	-	-	-	-	-	-	-
		Average Formant Frequency	-	2940	-	-	-	-	2682	-	-	-
		Standard Deviation	-	-	-	-	-	-	-	-	-	-

variation is from 1643 c/s for vowel ऍ (ऍ) to 4170 c/s for vowel e (ऐ) for males. In case of females third formant frequency is not clearly recorded in spectrograms and, therefore, it is not measured. For males average third formant frequency is from 2298 c/s for vowel ʌ (अ) to 3447 c/s for vowel i (इ). The individual fourth formant frequency range for males is from 2240 c/s for vowel O (ओ) to 5530 c/s for vowel e (ऐ) which does not include data of fourth formant frequency of vowel I (इ), U (उ), u (ऊ) . The variation of average fourth formant frequency is from 2867 c/s for vowel ʌ (अ) to 4865 c/s for vowel e (ऐ). The average fundamental and formant frequencies of males and females are also shown in **Table - 6**. These results are in similar to the results given by **PETERSON and BARNEY (23) for hVḍ and MAJUMDER, D. D., DUTTA, A. K. and GANGULI, N. R. (23)** for isolated vowels and **POLS, TROMP and PLOMP (27)** for hVṭ context.

The low values of the standard deviations, particularly for Hindi vowel (a), (U) and (u) in comparision to other vowels, indicate acoustic stability of these vowels. The phonetic identity of vowels seems to depend not on the absolute values of the formant frequencies but on the relative overall formant structure of the speaker. Third formant frequencies (F3) have larger standard deviation so it appears that this serves as an indicator to the formant structure of an individual speaker.

To study of context effect consonants on vowel formant frequency two pilot studies were done on CVC

TABLE – 6 : Average formant frequency of Males and Females

Words	AVERAGE FORMANT FREQUENCY (MALE)					AVERAGE FORMANT FREQUENCY (FEMALE)			
	F0	F1	F2	F3	F4	F0	F1	F2	F3
/hʌd/	154	715	1578	2298	2867	200	827	1543	-
/had/	152	606	1390	2417	3520	211	1138	2097	2940
/hɪd/	160	498	1440	2794	3290	227	587	1821	-
/hid/	166	569	1921	3447	4340	221	572	-	-
/hʊd/	169	493	1230	1966	3710	223	583	1950	-
/hud/	174	504	1251	2321	3780	252	561	1680	-
/hɛd/	153	624	1400	2382	3328	209	523	1587	2682
/hed/	168	650	2107	3157	4865	244	620	2372	-
/hɔd/	148	535	1270	2388	2987	198	477	1662	-
/hɔd/	172	689	1369	2394	3731	205	556	1537	-

and VCV syllables.

Table - 7, presents data of formant frequencies of 24 CVC nonsense syllables. The syllable had different initial and final consonants and were embeded with vowel /a/. The formant frequencies were measured at the centre of stationary state vowel parts of expanded broad band spectrogram (originally in the frequency range of 160Hz to 16KHz). The data of **Table - 7** is arranged as words with initial voicing consonants and words with final voicing consonants. In this table, for words with initial voicing consonants, individual first and second formant frequency in different consonant context varies from 565 c/s to 783 c/s and from 1090 c/s to 1360 c/s, respectively. The average frequency is found to be 650 c/s and 1226 c/s for first formant frequency and second formant frequency, respectively and standard deviation of F1 and F2 is 69.04 and 84.91, respectively. The formant frequencies of 8 words are higher of 6 words are below the average first formant frequency whereas for second formant, the formant frequencies of 6 words are above and of 8 words are below the average second formant frequency in the case of words with initial voicing consonants. For words with final voicing consonant, the first and second formant frequencies varies from 541 c/s to 695 c/s and 1131 c/s to 1347 c/s, respectively. While their average variation is measured to be 634 c/s and 1231 c/s, respectively for first and second formant frequencies. In this case, S. D. is 43.05 and 58.05, respectively. 8 words for first formant frequency

and 6 words for second formant frequency have the value higher than the average formant frequency whereas 6 words for first formant frequency and 8 words for second formant frequency have the value lower than average formant frequency.

Average value for first and second formant frequency for words with initial voiceless and words with final voiceless is also given in the same table and found to be 636 c/s and 1258 c/s, 652 c/s and 1253 c/s, respectively. Their S. D. is also measured as 56.33 and 59.10, 77.63 and 87.31, respectively.

From the **Table - 7**, it is also observed that for formant frequencies, S. D. are more when voiced consonants precedes than the case of voiced consonants follows it. Therefore, acoustical stability of vowels in final voicing consonants is more than in initial voicing consonants for Hindi vowels in CVC context. From the same table, it is also predicted that acoustical stability of vowels in initial voiceless consonants is more than in final voiceless consonants for Hindi vowels.

In **Table - 8**, CVC syllable arranged placewise to see the effect of place context on vowel characteristics. The consonants of same place are arranged together with their formant frequency shown. The average formant frequency of first and second formant frequency of Bilabial, Dental and Alveolar, Retroflex, Platal, Velar and Glottal are observed as 658 c/s and 1227 c/s, 655 c/s and 1274 c/s, 642 c/s and 1244 c/s, 614 c/s and 1223 c/s, 646 c/s and 1235 c/s,

TABLE—7: Formant Frequencies (CVC sounds) According to place of the Initial consonants.

WORDS WITH INITIAL VOICING	F1	F2	WORDS WITH FINAL VOICING	F1	F2
/bak/	565	1090	/dadʒʰ/	625	1208
/dadʒʰ/	625	1208	/gʰar/	606	1131
/gʰar/	606	1131	/tʃan/	565	1212
/datʃʰ/	783	1360	/dʰaj/	688	1333
/dʰaj/	688	1333	/bʰal/	653	1183
/bʰal/	653	1183	/Saw/	541	1166
/nah/	693	1306	/ham/	612	1183
/was/	565	1131	/ṭʰad/	646	1252
/radʰ/	686	1252	/tʰaḍ/	653	1224
/gapʰ/	653	1183	/radʰ/	686	1252
/maṭ/	734	1306	/ṭagʰ/	612	1265
/jaṭʰ/	565	1171	/ʃab /	646	1212
/dʒʰat/	565	1171	/tab/	653	1265
/dʒatʰ/	727	1333	/pʰag/	695	1347
Average	650	1226	Average	634	1231
Standard Deviation	69.04	84.91	Standard Deviation	43.05	58.05
INITIAL VOICELESS			FINAL VOICELESS		
Average	636	1258	Average	652	1253
Standard Deviation	56.33	59.10	Standard Deviation	77.63	87.31

TABLE – 8: Formant Frequencies (CVC sounds) According to place of the Initial consonants.

Place	Words	F1	F2
Bilabial (Front sounds)	/paʃ/	734	1306
	/mat̪/	734	1306
	/pʰag/	695	1347
	/bʰal/	653	1183
	/waS/	565	1131
	/bak/	565	1090
	Average	658	1227
	Standard Deviation	70.99	97.29
Dental and Alveolar (Middle Front)	/datʃʰ/	783	1360
	/nah/	693	1306
	/daj/	688	1333
	/tab/	653	1265
	/tʰad̪/	653	1224
	/lap/	571	1265
	/Saw/	541	1166
	Average	655	1274
	Standard Deviation	74.61	61.21
Retroflex (Middle sounds)	/radʰ/	686	1252
	/t̪ʰad/	646	1252
	/d̪adʒʰ/	625	1208
	/t̪agʰ/	612	1265
	Average	642	1244
	Standard Deviation	28.02	21.51
Platal (Middle back)	/dʒatʰ/	727	1333
	/ʃabʰ/	646	1212
	/dʒʰat/	565	1171
	/jat̪ʰ/	565	1171
	/tʃʰakʰ/	618	1237
	/tʃan/	565	1212
	Average	614	1223
	Standard Deviation	59.17	54.68
Velar and Glottal (Back sounds)	/kʰadʒ/	739	1391
	/gapʰ/	653	1183
	/katʃ/	622	1288
	/ham/	612	1183
	/gʰar/	606	1131
	Average	646	1235
	Standard Deviation	49.04	93.13

respectively. And their standard deviations are found as 70.99 and 97.29, 74.61 and 61.21, 28.02 and 21.51, 59.17 and 54.68, 49.04 and 93.13, respectively. It may be seen that there is no ordered effect on formant frequencies of Hindi vowels of consonants place of articulation. However, the acoustical stability of Hindi vowel in Retroflex sounds is most and in Bilabial (front sounds) sounds is least than among consonants place of articulation. Third formant frequency is not taken because it is not observed clearly in available spectrogram of all syllables.

Table - 9, presents data of formant frequencies of 24 VCV nonsense syllables having same initial and final vowels /a,i,u/. The 8 consonants used are / p, t, t, k, b, d, d, g / The formant frequency were measured at the centre of stationary state vowel parts of VCV syllables from its spectrograms taken from the computer facility of **STL, Stockholm, Sweden** having Sampling frequency 16.0 KHz, time window 10.0 ms, bandwidth 287.0 Hz and Gain varying from 6.0 db to 10.0 db. These sound samples were of a single speaker who is engaged in area of speech for more than 20 years. For individual voiced consonants sounds, the range of first formant frequency, second formant frequency and third formant frequency of vowel **a** is from 640 c/s to 720 c/s, 1120 c/s to 1200 c/s and 2080 c/s to 2400 c/s, respectively. For vowel **i** the range is 280 c/s to 300 c/s, 2320 c/s to 2440 c/s and 2960 c/s to 3040 c/s, respectively and for vowel **u** the range of voiced consonants is 280 c/s to 360 c/s, 720 c/s to

TABLE — 9: Formant Frequency of Vowels /a,i,u/ (VCV sounds) According to consonants placed of Articulation.

Words	F1	F2	F3	Words	F1	F2	F3	Words	F1	F2	F3
/ada/	680	1200	2240	/idi/	320	2320	3000	/udu/	320	800	2400
/ad^{h}a/	680	1200	2160	/id^{h}i/	300	2380	2980	/ud^{h}u/	320	860	2480
/ada/	720	1200	2340	/idi/	280	2400	2970	/udu/	320	880	2640
/ad^{h}a/	720	1200	2400	/id^{h}i/	310	2420	2980	/ud^{h}u/	320	840	2480
/aba/	680	1160	2160	/ibi/	320	2440	3000	/ubu/	340	800	2440
/ab^{h}a/	640	1200	2280	/ib^{h}i/	290	2400	2990	/ub^{h}u/	360	720	2520
/aga/	640	1200	2200	/igi/	280	2360	2980	/ugu/	320	840	2560
/ag^{h}a/	640	1160	2240	/ig^{h}i/	280	2360	3020	/ug^{h}u/	280	820	2480
/ama/	720	1200	208C	/imi/	320	2380	2990	/umu/	360	820	2480
/ana/	640	1140	2240	/ini/	330	2400	3000	/unu/	320	720	2480
/aja/	720	1200	2160	/iji/	300	2320	2990	/uju/	340	800	2400
/ara/	640	1120	2240	/iri/	290	2360	3000	/uru/	320	880	2400
/ala/	680	1120	2160	/ili/	300	2320	3040	/ulu/	320	880	2480
/awa/	680	1120	2240	/iwi/	280	2380	2960	-	-	-	-
Average	677	1172	2224	Average	300	2374	2992	Average	326	820	2480
Standard Deviation	31.94	33.47	78.62	Standard Deviation	16.90	35.79	19.43	Standard Deviation	19.82	51.44	64.68
				FOR VOICELESS CONSONANTS							
Average	707	1238	2294	Average	305	2362	2991	Average	310	822	2428
Stadnard	43.82	54.0	72.69	Standard	16.06	31.55	17.72	Standard	27.20	26.0	56.70

880 c/s and 2400 c/s to 2640 c/s for first, second and third formant frequency, respectively. The average formant frequency for voiced consonants are also shown in the same Table and is found to be 677 c/s, 1172 c/s, 2224 c/s for vowel a, and 300 c/s, 2374 c/s, 2992 c/s for vowel i, and 326 c/s, 820 c/s, 2480 c/s for vowel u, respectively. Standard deviations are also calculated and is found to be 31.94, 33.47, 78.62 for vowel **a**, 16.90, 35.79, 19.43 for vowel **i** and 19.82, 51.44, 64.68 for vowel **u**, respectively.

In the Table - 9, for the individual voiceless consonants, the average frequency is observed as 707 c/s, 1238 c/s, 2294 c/s for vowel **a**, 305 c/s, 2362 c/s, 2991 c/s for vowel **i** and 310 c/s, 822 c/s, 2428 c/s for vowel **u**, respectively. S. D. are also shown in the table and found as 43.82, 54.0, 72.69 for vowel **a**, 16.06, 31.55, 17.72 for vowel **i** and 27.20, 26.0, 56.70 for vowel **u**, respectively for voiceless consonants.

Average values of formant frequencies and S. D. of vowels /a,i,u/ (VCV sounds) in reference to consonants place of articulation is given in **Table - 10**. For bilabial sounds, the average value of first, second and third formant frequency is given as 693 c/s, 1180 c/s, 2247 c/s for vowel **a**, 308 c/s, 2390 c/s, 2990 c/s for vowel **i** and 248 c/s, 788 c/s, 2464 c/s for vowel **u**, respectively. Their S. D. are also determined as 29.81, 30.55, 104.34 for vowel **a**, 20.76, 25.16, 14.14 for vowel **i** and 16.0, 34.87, 40.97 for vowel **u**, respectively. The average value of first, second and third formant frequency for dental and alveolar sounds is obtained

TABLE — 10: Average values of Formant Frequencies and Standard Deviation of Vowels /a,i,u/ (VCV sounds) in reference to consonants place of Articulation.

Place	Vowels		F1	F2	F3
Bilabial (Front sounds)	a	Average	693	1180	2247
		S.D.	29.81	30.55	104.34
	i	Average	308	2390	2990
		S.D.	20.76	25.16	14.14
	u	Average	348	788	2464
		S.D.	16.0	34.87	40.79
Dental and Alveolar (Middle Front)	a	Average	690	1209	2311
		S.D.	34.22	72.39	79.89
	i	Average	307	2386	3001
		S.D.	14.84	31.55	21.78
	u	Average	303	834	2466
		S.D.	19.79	54.21	27.70
Retroflex (Middle sounds)	a	Average	680	1204	2212
		S.D.	25.29	57.13	34.87
	i	Average	303	2340	2982
		S.D.	12.30	25.29	18.33
	u	Average	324	836	2408
		S.D.	8.0	32.0	39.19
Velar and glottal (Back sounds)	a	Average	688	1208	2240
		S.D.	64.0	39.19	43.81
	i	Average	288	2360	2993
		S.D.	9.79	25.29	15.36
	u	Average	304	824	2464
		S.D.	19.59	14.96	78.38

as 690 c/s, 1209 c/s, 2311 c/s for vowel **a**, 307 c/s, 2386 c/s, 3001 c/s for vowel **i** and 303 c/s, 834 c/s, 2466 c/s for vowel **u**, respectively. Corresponding standard deviations are 34.22, 72.39, 79.89 for vowel **a**, 14.84, 31.55, 21.78 for vowel **i** and 19.79, 54.21, 27.70 for vowel **u**, respectively and are shown in the Table. Similar average value of retroflex sounds are found and they are 680 c/s, 1204 c/s, 2212 c/s for vowel **a**, 303 c/s, 2340 c/s, 2982 c/s for vowel **i** and 324 c/s, 836 c/s, 2408 c/s for vowel **u**, respectively, for first, second and third formant frequency. Their standard deviations are 25.29, 57.13, 34.87 for vowel **a**, 12.30, 25.29, 18.33 for vowel **i** and 8.0, 32.0, 39.19 for vowel **u** for first, second and third formant frequency, respectively. For velar and glottal sounds the average first, second and third formant frequency is observed as 688 c/s, 1208 c/s, 2240 c/s for vowel **a**, 288 c/s, 2360 c/s, 2993 c/s for vowel **i** and 304 c/s, 824 c/s, 2464 c/s for vowel **u**, respectively. Their corresponding standard deviations are 64.0, 39.19, 43.81 for vowel **a**, 9.79, 25.29, 15.36 for vowel **i** and 19.59, 14.96, 78.38 for vowel **u**, respectively. From Table, it is concluded that there is hardly any consonant effect on average values of formant frequencies and S. D. of vowels /a,i,u/ (for VCV sounds) in reference to consonants place of articulation.

As we have not found any consonant context effect on vowel formants in either CVC sounds or VCV sounds, we feel that vowel characteristics obtained hVd context may be taken as representative values of vowels.

There is considerable overlapping in F1 - F2 plane of Hindi vowels for males because of less % of intellegibility in comparision to **PETERSON and BARNEY (23) (Table - 2(b))** but for females the overlapping is negligible **(Fig. 2 and Fig. 3)** for Hindi vowels, which is well agreed with the results of **MAJUMDER, D. D., DUTTA, A. K. and GANGULI, N. R. (23)** who also found overlapping in F1 - F2 plane. It is may be due to the fact that Hindi has more flexilibity in its formant than English. There is also some overlapping in $F_3 - F_2$ plane (where $F_1^3 = F3 - F1$ and $F_2^3 = F3 - F2$) **(Fig. 4)** which hardly agreed with the results given by **MAJUMDER,. DUTTA and GANGULI (23)** because one reason may be that they have taken their speakers of well trained Hindi News announcers of AIR, New Delhi.

Therefore, there may be some possibility to obtained a clear seperation of different vowels and also clustering of the same vowels after giving proper weight to F3. Some spectrograms **(Fig. 5)** are taken from the computer facility of **STL, Stockholm, Sweden** for a single speaker of Hindi vowels /a,i,u/. The F1 - F2 plane of these Hindi vowels is shown in **Fig.- 6** and no overlapping is observed.

FUTURE PLAN:-

Digital signal processing has become a basic facility for speech research. The digitized speech wave - forms are processed by various technique, such as auto correlation, digital filtering, linear prediction, F F T etc. by programing on a general purpose computer which facilitates a great deal of speech analysis, synthesis and recognition.

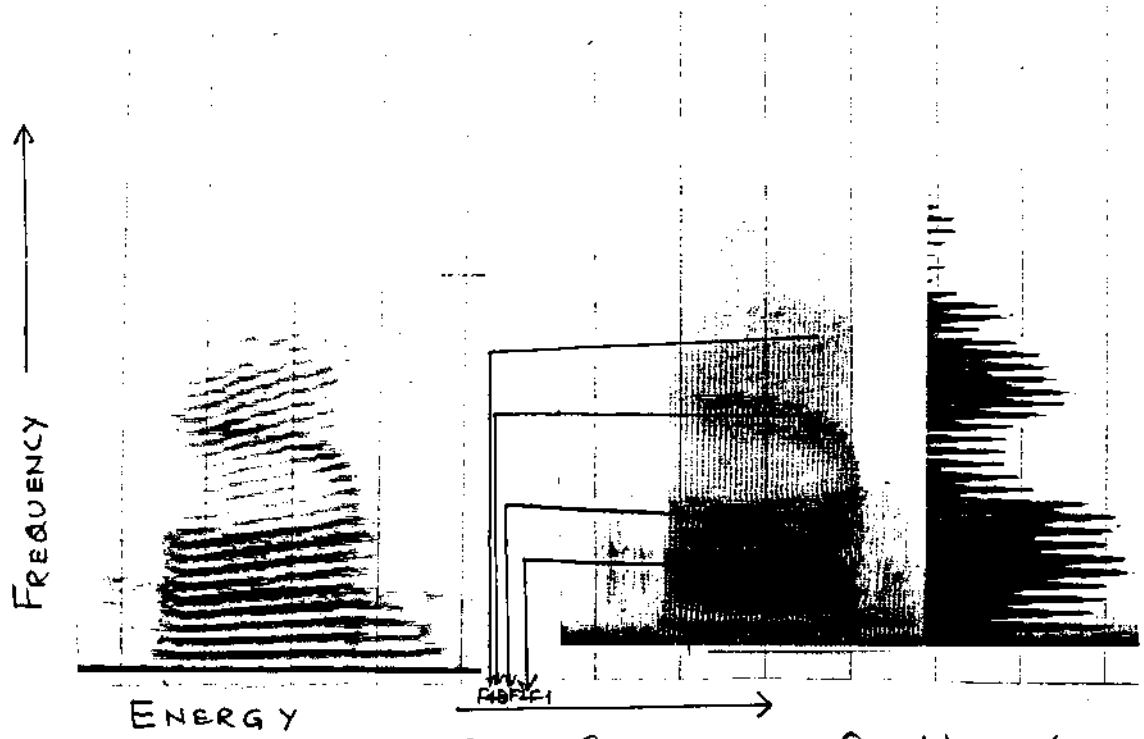


FIG-1:(a) NARROW BAND SPECTROGRAM OF WORD /had/

(b) BROAD BAND SPECTROGRAM OF WORD /had/

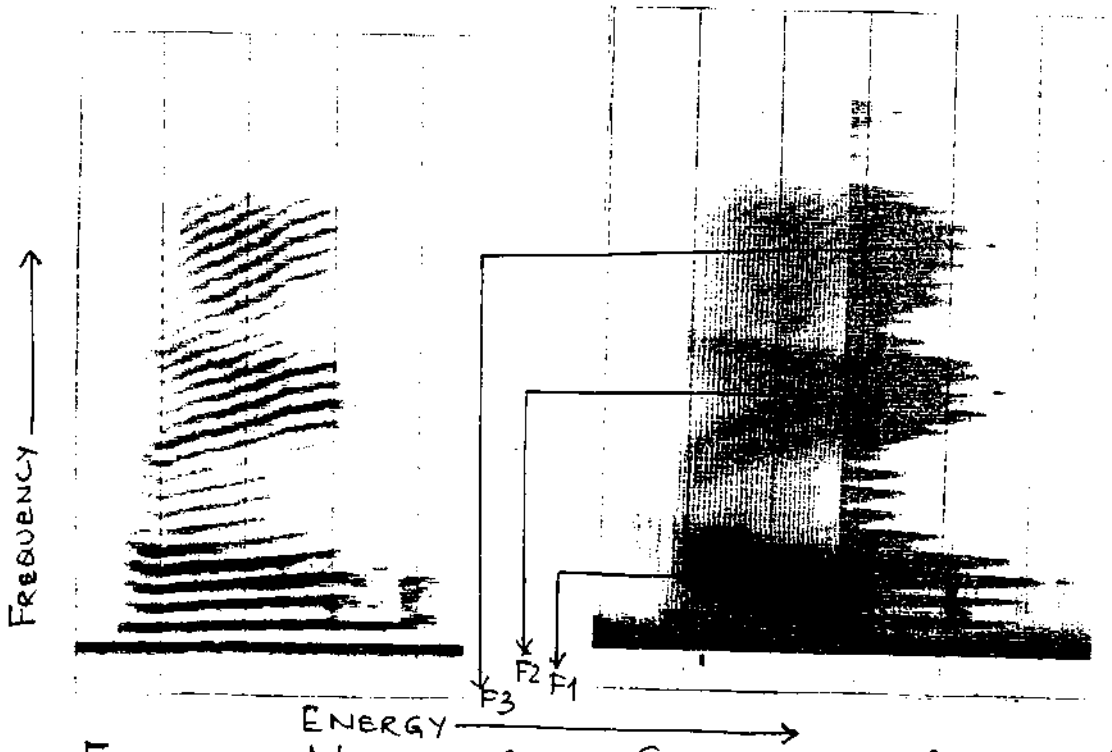


FIG-1:(c) NARROW BAND SPECTROGRAM OF WORD /hed/

(d) BROAD BAND SPECTROGRAM OF WORD /hed/

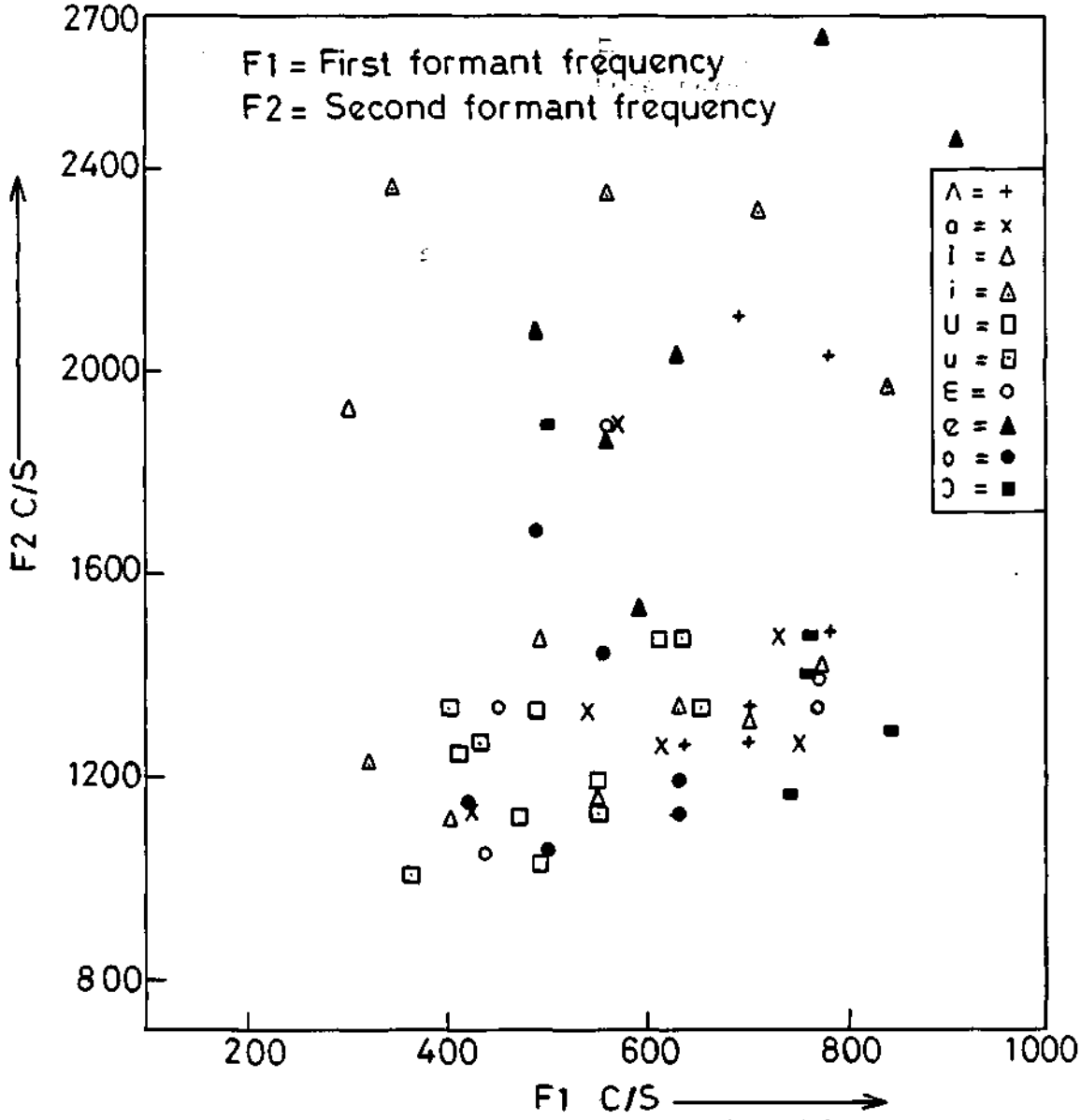


Fig.2 Distribution of Hindi vowel in F1-F2 plane (Males only)

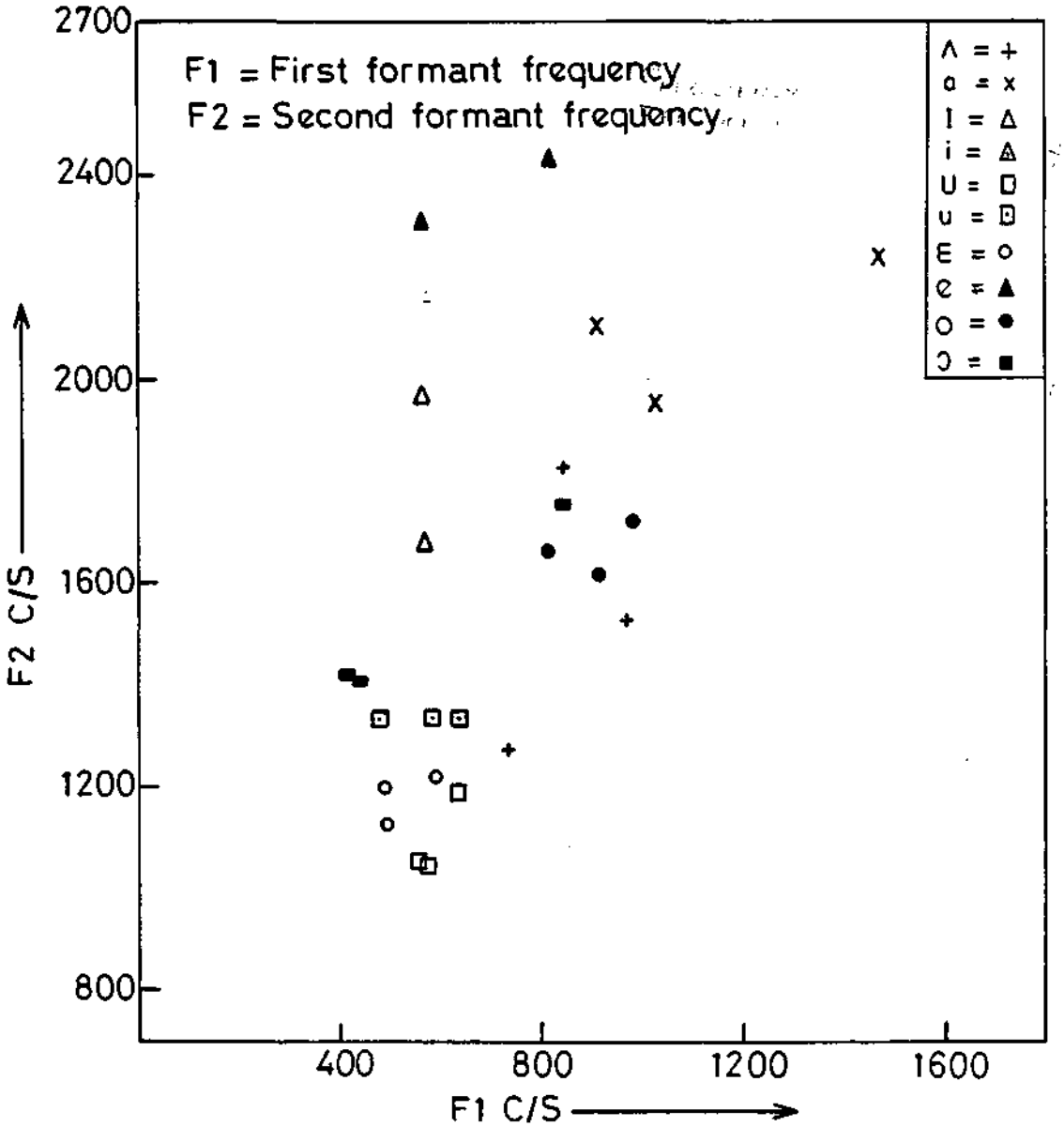


Fig.3 Distribution of Hindi vowels in F1-F2 plane (Females only)

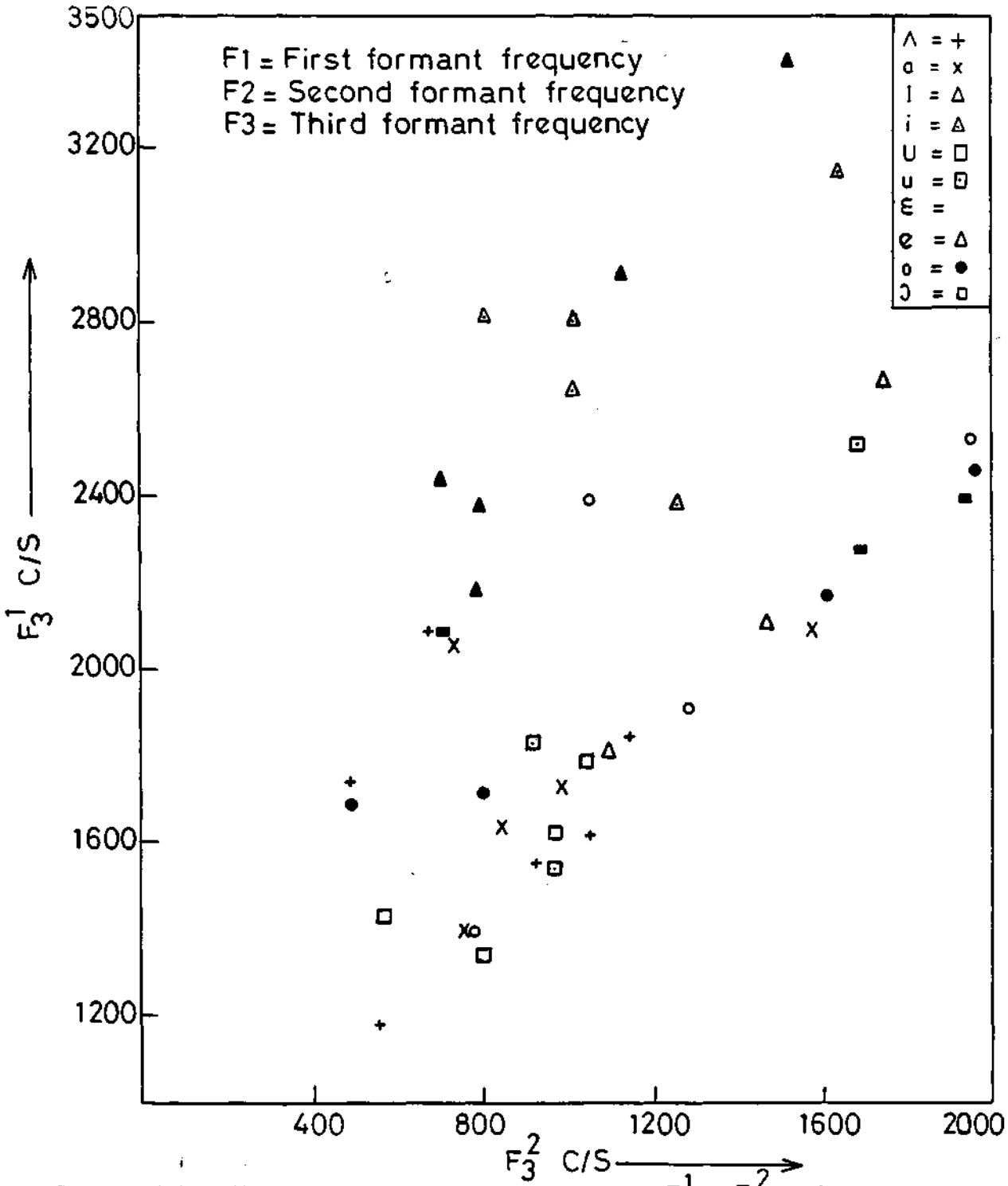


Fig. 4 Distribution of Hindi vowels in $F_3^1 - F_3^2$ plane (Males only)

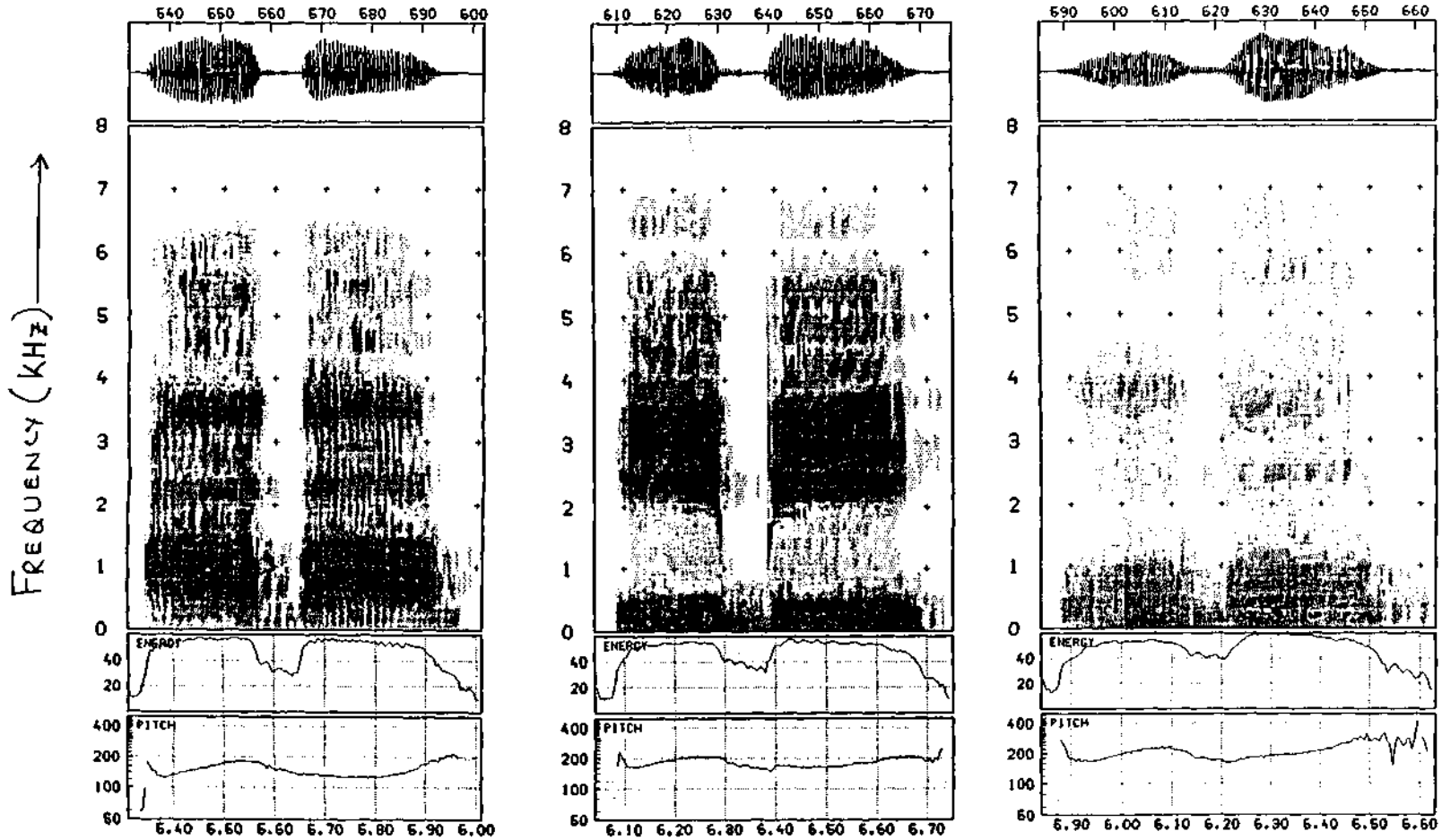


FIG-6: BROAD BAND SPECTROGRAM

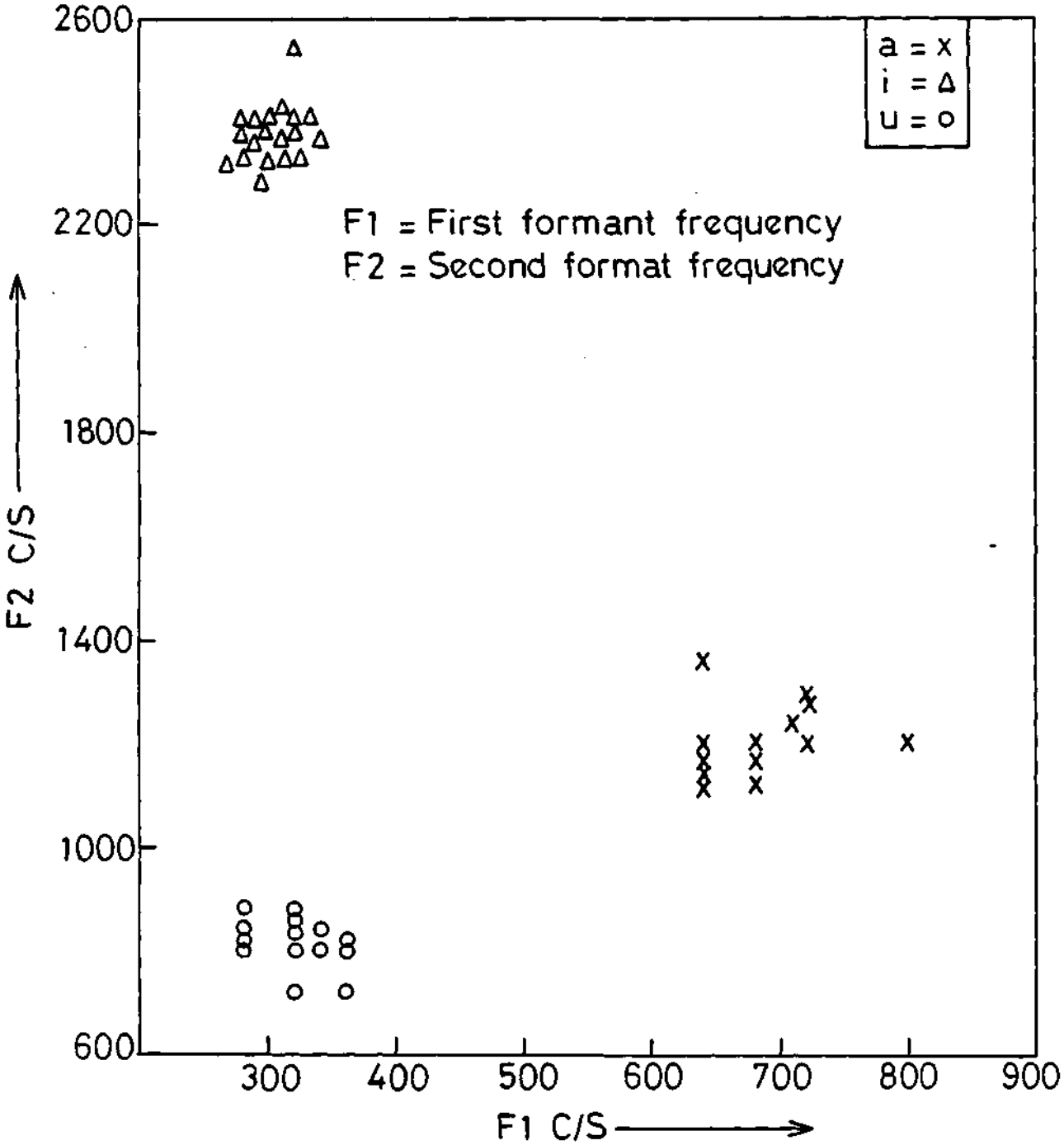


Fig.6: Distribution of Hindi vowels in F1-F2 plane (Single Speaker)

For digitization speech signal A/D and D/A system are going to be interfaced with the University computer. I shall study, evaluate and test softwares available with the new system for effective and versatile application to our problems. Software programs available in digital signal processing literature will be implemented on our computer. Soon, LPA11-K, AD11-K and AA11-K is giong to be interfaced with the VAX-11/780 computer system. Recently ordered have given to ILS PC ful with A/D and D/A cards of Data Translation (USA) for our research work.

We analyse natural speech using conventional sound spectrograph and further analysis is made by computer to determine the various acoustic parameters of speech such as fundamental frequency formants, their amplitude and width, voiced / un-voiced detection, segmentation of connected speech, etc. We are trying to develop editing facilities which will enables us to prepare stimuli for perception experiments.

The major problems of speech research is to extract the information bearing parameters of speech sounds as there is a lot of redundency in the signal. In fact, these information bearing elements convey the message from speaker to listeners. These can be determined through perceptual experiments by analysis - synthesis technique in that, firstly the parameters are determined by analysing natural speech signals and then by speech synthesis.

I have to survey pertinent literature to be

aware of the latest state - of -art in passing out my experiments.

Chapter 5

B I B L I O G R A P H Y

1. STEVENS, K. N. & HOUSE, A. S. "Development of a Quantitative Description of Vowel Articulation" ,JASA, Vol., 27, 1955; pp. 489 - 493.

2. POLS, L. C. W., KAMP, & PLOMP, R. "Perception and Physical Space of Vowel Sounds", JASA, Vol., 46, 1969; pp. 458 - 467.

3. HIRSH, I. J. "Auditory Perception of Temporal Order" JASA, Vol., 31, 1957; pp. 759 - 767.

4. KLEIN, W., PLOMP, R. & POLS, L. C. W. "Vowel Spectra, Vowel Spaces, and Vowel Identification", JASA, Vol., 48, 197Ø; pp. 999 - 1ØØ9.

5. SCHOUTEN et al. "Pitch of The Residue", JASA, Vol., 34, 1962, pp. 1418 - 1424.

6. RUPF, J. A., HUGES & HOUSE "Effect of Interaural Switching On The Recognition of Speech Sounds", JASA, Vol., 51, 1972.

7. FANT, C. G. M. "Nonuniform Vowel Normalization", R. Inst. Tech., Speech Trans. Lab., Stockholm, Sweden, Q. Prog. Stat. Rep., October (1976), pp. 1 - 9.

8. SCOTT, B. L. "Temporal Factors in Vowel Perception", JASA, Vol., 6Ø, 1976; pp. 1354 - 1365.

9. DIEHL, R. L., MCCUSKER S. B. & CHAPMAN, L. S. "Perceiving Vowels in Isolation and in Consonant Context", JASA, Vol., 69, 1981.

1Ø. TRAUNMIILLER, H. "Perceptual Dimension of Openness in Vowels" JASA, Vol., 69, 1981.

11. BERNSTEIN, J. "Formant - Based Representation of Auditory Similarity Among Vowel - Like Sounds", JASA, Vol., 69, No. 4; 1981.

12. KEWLEY - PORT, D. "Measurement of Formant Transitions in Naturally Produced Stop Consonant - Vowel Syllables", JASA, Vol., 72, No. 2; 1982.

13. STRANGE, W., JENKINS, J. J., JOHNSON, T. L. "Dynamic Specification of Coarticulated Vowels", JASA, Vol., 74, No. 3; 1983.

14. RAKERD, B., VERBRUGGE, R. R., SHANKWEILER, D. P. "Monitoring For Vowels in Isolation and in a Consonantal Context", JASA, VØl., 76, No. 1; 1984.

15. GOTTFRIED, T. L. "Intelligibility of Vowels Sung By a Countertenor", JASA, Vol., 79, No. 1; 1986.

16. HOWELL, P. & VAUSE, L. " Acoustical Analysis and Perception Of Vowels in Stuttered Speech", JASA, Vol., 79, No. 5; 1986.

17. BROAD, D. J. " Piecewise - Planar Vowel Formant Distribution Across Speakers ", JASA, Vol., 69, No. 5, 1981; pp. 1423 - 1429.

18. PETERSON, G. E. & LEHISTE, I. " Duration Of Syllable Nuclei in English ", JASA, Vol., 32, 196Ø; pp. 693 - 7Ø3.

19. HUGGINS, A. W. F. "Just Noticeable Differences For Segment Duration in Natural Speech", JASA, Vol., 51, No. 4; 1972.

2Ø. SOLI, S. D. "Structure and Duration of Vowels Togather Specify Fricative Voicing", JASA, Vol., 72, No. 2; 1982.

21. BLADON, R. A. W. "Modeling The Judgement of Vowel Quality

Difference", JASA, Vol., 69, No. 5, 1981; pp. 1414 - 1422.

22. COWAN, N. "The Use of Auditory and Phonetic Momery in Vowel Discrimination", JASA, Vol., 79, No. 2; 1986.

23. FANT, C. G. M. "On The Prediction of Formamt Levels and Spectrum Envelopes From Formant Frequencies For Roman Jakobson (The Hague: Mounton,1956) pp. 109 - 120.

PETERSON, G. E. & BARNEY, H. L. "Control Methods Used in Study of Vowels", JASA, Vol., 24, No. 2, 1952; pp. 175 - 184.

MAJUMDER, D. D., DUTTA, A. K. & GANGULI, N. R. "Some Studies on Acoustic Features of Human Speech in Relation to Hindi Speech Sounds" Indian J. of Physics, 47, 1973;pp. 598-613.

24. FLANAGAN, J. L. "Difference Limen For Formant Amplitude", J. Speech and Hearing Research, Vol., 22, 1957; pp. 205 - 212.

25. STEVENS, K. N. & HOUSE, A. S. "An Acoustical Theory of Vowel Production and Some of Its Implications", J. Speech and Hearing Research, Vol., 4, 1961; pp. 303 - 320.

26. PLOMP, POLS & GEER "Dimensional Analysis Of Vowel Spectra", JASA; 1967.

28. FANT, C. G. M. "Speech Sounds And Features", (MIT, Cambridge, MA.); 1973.

27. POLS, L. C. W., TROMP, H. R. C. & PLOMP, R. "Frequency Analysis of Dutch Vowels From 50 Male Speakers", JASA, Vol., 53, No. 4; 1973, pp. 1093 - 1101.

29. BENNETT, S. "Vowel Formant Frequency Characteristics of Preadolescent Males and Females", JASA, Vol., 69, No. 1; 1981.

30. SCHEFFERS, M. T. M. "Discrimination of Fundamental Frequency of Synthesized Vowel Sounds in a Noise Back - ground", JASA, Vol., 76, No. 2; 1984.